****

MPSRCH (TM)

****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Mon Aug 3 13:21:28 1998; MasPar time 5366.00 Seconds
1374.553 Million cell updates/sec

Tabular output not generated.

Title: >US-08-912-951-1
Description: (1-4015) from US08912951.seq
Perfect Score: 4015
N.A. Sequence: 1 GCAGCGCTGCGTCCTGCTGC.....TTTTTCAGTTTTGAAAAAAA 4015
Comp: CGTCGCGACGCAGGACGACG.....AAAAAGTCAAAACTTTTTTT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 532261 seqs, 918536377 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: embl55
1:em_ba 2:em_htg 3:em_hum1 4:em_hum2 5:em_in 6:em_om
7:em_or 8:em_ov 9:em_pat 10:em_pl 11:em_ro
Database: genbank107
12:gb_ba 13:gb_htg 14:gb_in 15:gb_om 16:gb_ov 17:gb_pat
18:gb_ph 19:gb_pl 20:gb_pri 21:gb_pr2 22:gb_ro 23:gb_st
24:gb_sts 25:gb_sy 26:gb_un 27:gb_vi

Statistics: Mean 12.056; Variance 6.037; scale 1.997

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	4015	100.0	4015	21	AF015950	Homo sapiens telomeras	0.00e+00
	2	4001	99.7	4027	21	AF018167	Homo sapiens telomeras	0.00e+00
	3	3439	85.7	8960	25	AF043739	Synthetic construct hu	0.00e+00
	4	920	22.9	3426	22	AF051911	Mus musculus telomeras	0.00e+00
	5	129	3.2	201	22	AF029235	Mus musculus telomeras	2.00e-73
	6	67	1.7	7218	17	I66494	Sequence 14 from paten	3.56e-26
c	7	49	1.2	7218	17	I66494	Sequence 14 from paten	1.15e-13
	8	42	1.0	215	17	I28278	Sequence 5 from patent	4.08e-09
c	9	37	0.9	215	17	I28278	Sequence 5 from patent	5.00e-06
	10	36	0.9	565	17	E04076	gDNA encoding envelope	1.98e-05
	11	30	0.7	201	17	A10161	Synthetic DNA for prep	5.17e-02
	12	30	0.7	201	17	A10158	S.griseus gene for pre	5.17e-02
c	13	30	0.7	201	17	A10162	Synthetic DNA for prep	5.17e-02
c	14	30	0.7	201	17	A10159	S.griseus gene for pre	5.17e-02
c	15	28	0.7	565	17	E04076	gDNA encoding envelope	5.96e-01
c	16	27	0.7	904	27	HS4BHLF	Epstein-Barr virus Bam	1.94e+00
c	17	27	0.7	1572	15	BTU21091	Bos taurus DNA polymer	1.94e+00
	18	29	0.7	1663	15	MVU92534	Mustela vison microsat	1.78e-01
	19	27	0.7	4053	14	DRO28DC16Z	Drosophila melanogaste	1.94e+00
	20	27	0.7	5050	27	HS4U2IR2A	Epstein-Barr virus (B9	1.94e+00
	21	27	0.7	5400	12	SCACTVA	S.coelicolor 6 actVA r	1.94e+00
	22	29	0.7	6011	27	HS4RSIRZ2T	Epstein-Barr virus (wi	1.78e-01
	23	29	0.7	15807	12	MEU72662	Methylobacterium extor	1.78e-01
c	24	27	0.7	36819	12	MTY20H10	Mycobacterium tubercul	1.94e+00
c	25	28	0.7	39073	19	SPAC2E11	S.pombe chromosome I c	5.96e-01
	26	27	0.7	64078	14	AC004365	Drosophila melanogaste	1.94e+00
	27	29	0.7	76094	21	HS179D3A	Human DNA sequence fro	1.78e-01
c	28	27	0.7	81023	14	AC001659	Drosophila melanogaste	1.94e+00
	29	27	0.7	172281	27	EBV	Epstein-Barr virus (EB	1.94e+00
	30	27	0.7	184113	27	HS4B958RAJ	Epstein-Barr virus, ar	1.94e+00
	31	26	0.6	1341	16	CHKD1LRB	Gallus domesticus D1-l	6.15e+00
	32	25	0.6	2842	12	S65123	lip=extracellular lipa	1.89e+01
	33	26	0.6	2926	12	SMPROBAG	S.marcescens proBA gen	6.15e+00
	34	26	0.6	2926	12	SMAPROBA	S.marcescens proBA ope	6.15e+00
	35	26	0.6	5591	12	AF027499	Azotobacter vinelandii	6.15e+00
	36	26	0.6	12949	12	AF011922	Pseudomonas aeruginosa	6.15e+00
	37	25	0.6	30000	27	BHV130KB	Bovine herpesvirus typ	1.89e+01
c	38	26	0.6	43034	21	HSN5H6	Human DNA sequence fro	6.15e+00
c	39	26	0.6	49634	21	AC002398	Human DNA from chromos	6.15e+00
c	40	26	0.6	85132	21	U82668	Homo sapiens shox gene	6.15e+00
	41	25	0.6	131530	21	HUAC002303	Homo sapiens Chromosom	1.89e+01
c	42	25	0.6	190289	27	MCU60315	Molluscum contagiosum	1.89e+01
	43	25	0.6	224681	13	HS44N10	Human DNA sequence ***	1.89e+01
	44	26	0.6	247731	13	CEY79H2	Caenorhabditis elegans	6.15e+00
	45	26	0.6	342884	13	CEY75B8	Caenorhabditis elegans	6.15e+00

ALIGNMENTS

RESULT 1
LOCUS AF015950 4015 bp mRNA PRI 16-AUG-1997
DEFINITION Homo sapiens telomerase reverse transcriptase (hTRT) mRNA, complete cds.
ACCESSION AF015950
NID g2330016
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4015)
AUTHORS Nakamura,T.M., Morin,G.B., Chapman,K.B., Weinrich,S.L., Andrews,W.H., Lingner,J., Harley,C.B. and Cech,T.R.
TITLE Telomerase catalytic subunit homologs from fission yeast and human
JOURNAL Science 277 (5328), 955-959 (1997)
MEDLINE 97400623
REFERENCE 2 (bases 1 to 4015)
AUTHORS Morin,G.B.
TITLE Direct Submission
JOURNAL Submitted (24-JUL-1997) Geron Corporation, 230 Constitution Drive, Menlo Park, CA 94025, USA
FEATURES Location/Qualifiers
source 1. .4015
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="kidney"
/map="5p"
/dev_stage="embryo"
/chromosome="5"
gene 1. .4015
/gene="hTRT"
CDS 56. .3454
/gene="hTRT"
/codon_start=1
/product="telomerase reverse transcriptase"
/db_xref="PID:g2330017"
/translation="MPRAPRCRAVRSLLRSHYREVLPLATFVRRLGPQGWRLVQRGDP

15 Aug 97

```
                  AAFRALVAQCLVCVPWDARPPPAAPSFRQVSCLKELVARVLQRLCERGAKNVLAFGFA
                  LLDGARGGPPEAFTTSVRSYLPNTVTDALRGSGAWGLLLRRVGDDVLVHLLARCALFV
                  LVAPSCAYQVCGPPLYQLGAATQARPPPHASGPRRRLGCERAWNHSVREAGVPLGLPA
                  PGARRRGGSASRSLPLPKRPRRGAAPEPERTPVGQGSWAHPGRTRGPSDRGFCVVSPA
                  RPAEEATSLEGALSGTRHSHPSVGRQHHAGPPSTSRPPRPWDTPCPPVYAETKHFLYS
                  SGDKEQLRPSFLLSSLRPSLTGARRLVETIFLGSRPWMPGTPRRLPRLPQRYWQMRPL
                  FLELLGNHAQCPYGVLLKTHCPLRAAVTPAAGVCAREKPQGSVAAPEEEDTDPRRLVQ
                  LLRQHSSPWQVYGFVRACLRRLVPPGLWGSRHNERRFLRNTKKFISLGKHAKLSLQEL
                  TWKMSVRDCAWLRRSPGVGCVPAAEHRLREEILAKFLHWLMSVYVVELLRSFFYVTET
                  TFQKNRLFFYRKSVWSKLQSIGIRQHLKRVQLRELSEAEVRQHREARPALLTSRLRFI
                  PKPDGLRPIVNMDYVVGARTFRREKRAERLTSRVKALFSVLNYERARRPGLLGASVLG
                  LDDIHRAWRTFVLRVRAQDPPPELYFVKVDVTGAYDTIPQDRLTEVIASIIKPQNTYC
                  VRRYAVVQKAAHGHVRKAFKSHVSTLTDLQPYMRQFVAHLQETSPLRDAVVIEQSSSL
                  NEASSGLFDVFLRFMCHHAVRIRGKSYVQCQGIPQGSILSTLLCSLCYGDMENKLFAG
                  IRRDGLLLRLVDDFLLVTPHLTHAKTFLRTLVRGVPEYGCVVNLRKTVVNFPVEDEAL
                  GGTAFVQMPAHGLFPWCGLLLDTRTLEVQSDYSSYARTSIRASLTFNRGFKAGRNMRR
                  KLFGVLRLKCHSLFLDLQVNSLQTVCTNIYKILLLQAYRFHACVLQLPFHQQVWKNPT
                  FFLRVISDTASLCYSILKAKNAGMSLGAKGAAGPLPSEAVQWLCHQAFLLKLTRHRVT
                  YVPLLGSLRTAQTQLSRKLPGTTLTALEAAANPALPSDFKTILD"
  COUNT      663 a   1363 c   1275 g    714 t
GIN

  Query Match          100.0%;  Score 4015;  DB 21;  Length 4015;
  Best Local Similarity 100.0%;  Pred. No. 0.00e+00;
  Matches  4015;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 GCAGCGCTGCGTCCTGCTGCGCACGTGGGAAGCCCTGGCCCCGGCCACCCCCGCGATGCC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1 GCAGCGCTGCGTCCTGCTGCGCACGTGGGAAGCCCTGGCCCCGGCCACCCCCGCGATGCC 60

Db     61 GCGCGCTCCCCGCTGCCGAGCCGTGCGCTCCCTGCTGCGCAGCCACTACCGCGAGGTGCT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     61 GCGCGCTCCCCGCTGCCGAGCCGTGCGCTCCCTGCTGCGCAGCCACTACCGCGAGGTGCT 120

Db    121 GCCGCTGGCCACGTTCGTGCGGCGCCTGGGGCCCCAGGGCTGGCGGCTGGTGCAGCGCGG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    121 GCCGCTGGCCACGTTCGTGCGGCGCCTGGGGCCCCAGGGCTGGCGGCTGGTGCAGCGCGG 180

Db    181 GGACCCGGCGGCTTTCCGCGCGCTGGTGGCCCAGTGCCTGGTGTGCGTGCCCTGGGACGC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    181 GGACCCGGCGGCTTTCCGCGCGCTGGTGGCCCAGTGCCTGGTGTGCGTGCCCTGGGACGC 240

Db    241 ACGGCCGCCCCCCCGCCGCCCCCTCCTTCCGCCAGGTGTCCTGCCTGAAGGAGCTGGTGGC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    241 ACGGCCGCCCCCCGCCGCCCCCTCCTTCCGCCAGGTGTCCTGCCTGAAGGAGCTGGTGGC 300

Db    301 CCGAGTGCTGCAGAGGCTGTGCGAGCGCGGCGCGAAGAACGTGCTGGCCTTCGGCTTCGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    301 CCGAGTGCTGCAGAGGCTGTGCGAGCGCGGCGCGAAGAACGTGCTGGCCTTCGGCTTCGC 360

Db    361 GCTGCTGGACGGGGCCCGCGGGGGCCCCCCCCGAGGCCTTCACCACCAGCGTGCGCAGCTA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    361 GCTGCTGGACGGGGCCCGCGGGGGCCCCCCCCGAGGCCTTCACCACCAGCGTGCGCAGCTA 420

Db    421 CCTGCCCAACACGGTGACCGACGCACTGCGGGGGAGCGGGGCGTGGGGGCTGCTGCTGCG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    421 CCTGCCCAACACGGTGACCGACGCACTGCGGGGGAGCGGGGCGTGGGGGCTGCTGCTGCG 480

Db    481 CCGCGTGGGCGACGACGTGCTGGTTCACCTGCTGGCACGCTGCGCGCTCTTTGTGCTGGT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    481 CCGCGTGGGCGACGACGTGCTGGTTCACCTGCTGGCACGCTGCGCGCTCTTTGTGCTGGT 540

Db    541 GGCTCCCAGCTGCGCCTACCAGGTGTGCGGGCCGCCGCTGTACCAGCTCGGCGCTGCCAC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    541 GGCTCCCAGCTGCGCCTACCAGGTGTGCGGGCCGCCGCTGTACCAGCTCGGCGCTGCCAC 600

Db    601 TCAGGCCCGGCCCCCGCCACACGCTAGTGGACCCCGAAGGCGTCTGGGATGCGAACGGGC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    601 TCAGGCCCGGCCCCCGCCACACGCTAGTGGACCCCGAAGGCGTCTGGGATGCGAACGGGC 660

Db    661 CTGGAACCATAGCGTCAGGGAGGCCGGGGTCCCCCTGGGCCTGCCAGCCCCGGGTGCGAG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    661 CTGGAACCATAGCGTCAGGGAGGCCGGGGTCCCCCTGGGCCTGCCAGCCCCGGGTGCGAG 720

Db    721 GAGGCGCGGGGGCAGTGCCAGCCGAAGTCTGCCGTTGCCCAAGAGGCCCAGGCGTGGCGC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    721 GAGGCGCGGGGGCAGTGCCAGCCGAAGTCTGCCGTTGCCCAAGAGGCCCAGGCGTGGCGC 780

Db    781 TGCCCCTGAGCCGGAGCGGACGCCCGTTGGGCAGGGGTCCTGGGCCCACCCGGGCAGGAC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    781 TGCCCCTGAGCCGGAGCGGACGCCCGTTGGGCAGGGGTCCTGGGCCCACCCGGGCAGGAC 840

Db    841 GCGTGGACCGAGTGACCGTGGTTTCTGTGTGGTGTCACCTGCCAGACCCGCCGAAGAAGC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    841 GCGTGGACCGAGTGACCGTGGTTTCTGTGTGGTGTCACCTGCCAGACCCGCCGAAGAAGC 900

Db    901 CACCTCTTTGGAGGGTGCGCTCTCTGGCACGCGCCACTCCCACCCATCCGTGGGCCGCCA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    901 CACCTCTTTGGAGGGTGCGCTCTCTGGCACGCGCCACTCCCACCCATCCGTGGGCCGCCA 960

Db    961 GCACCACGCGGGCCCCCCATCCACATCGCGGCCACCACGTCCCTGGGACACGCCTTGTCC 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    961 GCACCACGCGGGCCCCCCATCCACATCGCGGCCACCACGTCCCTGGGACACGCCTTGTCC 1020

Db   1021 CCCGGTGTACGCCGAGACCAAGCACTTCCTCTACTCCTCAGGCGACAAGGAGCAGCTGCG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1021 CCCGGTGTACGCCGAGACCAAGCACTTCCTCTACTCCTCAGGCGACAAGGAGCAGCTGCG 1080

Db   1081 GCCCTCCTTCCTACTCAGCTCTCTGAGGCCCAGCCTGACTGGCGCTCGGAGGCTCGTGGA 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1081 GCCCTCCTTCCTACTCAGCTCTCTGAGGCCCAGCCTGACTGGCGCTCGGAGGCTCGTGGA 1140

Db   1141 GACCATCTTTCTGGGTTCCAGGCCCTGGATGCCAGGGACTCCCCGCAGGTTGCCCCGCCT 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1141 GACCATCTTTCTGGGTTCCAGGCCCTGGATGCCAGGGACTCCCCGCAGGTTGCCCCGCCT 1200

Db   1201 GCCCCAGCGCTACTGGCAAATGCGGCCCCTGTTTCTGGAGCTGCTTGGGAACCACGCGCA 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1201 GCCCCAGCGCTACTGGCAAATGCGGCCCCTGTTTCTGGAGCTGCTTGGGAACCACGCGCA 1260

Db   1261 GTGCCCCTACGGGGTGCTCCTCAAGACGCACTGCCCGCTGCGAGCTGCGGTCACCCCAGC 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1261 GTGCCCCTACGGGGTGCTCCTCAAGACGCACTGCCCGCTGCGAGCTGCGGTCACCCCAGC 1320

Db   1321 AGCCGGTGTCTGTGCCCGGGAGAAGCCCCAGGGCTCTGTGGCGGCCCCCGAGGAGGAGGA 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1321 AGCCGGTGTCTGTGCCCGGGAGAAGCCCCAGGGCTCTGTGGCGGCCCCCGAGGAGGAGGA 1380

Db   1381 CACAGACCCCCGTCGCCTGGTGCAGCTGCTCCGCCAGCACAGCAGCCCCTGGCAGGTGTA 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1381 CACAGACCCCCGTCGCCTGGTGCAGCTGCTCCGCCAGCACAGCAGCCCCTGGCAGGTGTA 1440

Db   1441 CGGCTTCGTGCGGGCCTGCCTGCGCCGGCTGGTGCCCCCAGGCCTCTGGGGCTCCAGGCA 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1441 CGGCTTCGTGCGGGCCTGCCTGCGCCGGCTGGTGCCCCCAGGCCTCTGGGGCTCCAGGCA 1500

Db   1501 CAACGAACGCCGCTTCCTCAGGAACACCAAGAAGTTCATCTCCCTGGGGAAGCATGCCAA 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1501 CAACGAACGCCGCTTCCTCAGGAACACCAAGAAGTTCATCTCCCTGGGGAAGCATGCCAA 1560

Db   1561 GCTCTCGCTGCAGGAGCTGACGTGGAAGATGAGCGTGCGGGACTGCGCTTGGCTGCGCAG 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1561 GCTCTCGCTGCAGGAGCTGACGTGGAAGATGAGCGTGCGGGACTGCGCTTGGCTGCGCAG 1620

Db   1621 GAGCCCAGGGGTTGGCTGTGTTCCGGCCGCAGAGCACCGTCTGCGTGAGGAGATCCTGGC 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1621 GAGCCCAGGGGTTGGCTGTGTTCCGGCCGCAGAGCACCGTCTGCGTGAGGAGATCCTGGC 1680

Db   1681 CAAGTTCCTGCACTGGCTGATGAGTGTGTACGTCGTCGAGCTGCTCAGGTCTTTCTTTTA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1681 CAAGTTCCTGCACTGGCTGATGAGTGTGTACGTCGTCGAGCTGCTCAGGTCTTTCTTTTA 1740

Db   1741 TGTCACGGAGACCACGTTTCAAAAGAACAGGCTCTTTTTCTACCGGAAGAGTGTCTGGAG 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1741 TGTCACGGAGACCACGTTTCAAAAGAACAGGCTCTTTTTCTACCGGAAGAGTGTCTGGAG 1800
```

```
Db  1801 CAAGTTGCAAAGCATTGGAATCAGACAGCACTTGAAGAGGGTGCAGCTGCGGGAGCTGTC 1860
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1801 CAAGTTGCAAAGCATTGGAATCAGACAGCACTTGAAGAGGGTGCAGCTGCGGGAGCTGTC 1860

Db  1861 GGAAGCAGAGGTCAGGCAGCATCGGGAAGCCAGGCCCGCCCTGCTGACGTCCAGACTCCG 1920
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1861 GGAAGCAGAGGTCAGGCAGCATCGGGAAGCCAGGCCCGCCCTGCTGACGTCCAGACTCCG 1920

Db  1921 CTTCATCCCCAAGCCTGACGGGCTGCGGCCGATTGTGAACATGGACTACGTCGTGGGAGC 1980
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1921 CTTCATCCCCAAGCCTGACGGGCTGCGGCCGATTGTGAACATGGACTACGTCGTGGGAGC 1980

Db  1981 CAGAACGTTCCGCAGAGAAAAGAGGGCCGAGCGTCTCACCTCGAGGGTGAAGGCACTGTT 2040
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1981 CAGAACGTTCCGCAGAGAAAAGAGGGCCGAGCGTCTCACCTCGAGGGTGAAGGCACTGTT 2040

Db  2041 CAGCGTGCTCAACTACGAGCGGGCGCGGCGCCCCGGCCTCCTGGGCGCCTCTGTGCTGGG 2100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2041 CAGCGTGCTCAACTACGAGCGGGCGCGGCGCCCCGGCCTCCTGGGCGCCTCTGTGCTGGG 2100

Db  2101 CCTGGACGATATCCACAGGGCCTGGCGCACCTTCGTGCTGCGTGTGCGGGCCCAGGACCC 2160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2101 CCTGGACGATATCCACAGGGCCTGGCGCACCTTCGTGCTGCGTGTGCGGGCCCAGGACCC 2160

Db  2161 GCCGCCTGAGCTGTACTTTGTCAAGGTGGATGTGACGGGCGCGTACGACACCATCCCCCA 2220
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2161 GCCGCCTGAGCTGTACTTTGTCAAGGTGGATGTGACGGGCGCGTACGACACCATCCCCCA 2220

Db  2221 GGACAGGCTCACGGAGGTCATCGCCAGCATCATCAAACCCCAGAACACGTACTGCGTGCG 2280
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2221 GGACAGGCTCACGGAGGTCATCGCCAGCATCATCAAACCCCAGAACACGTACTGCGTGCG 2280

Db  2281 TCGGTATGCCGTGGTCCAGAAGGCCGCCCATGGGCACGTCCGCAAGGCCTTCAAGAGCCA 2340
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2281 TCGGTATGCCGTGGTCCAGAAGGCCGCCCATGGGCACGTCCGCAAGGCCTTCAAGAGCCA 2340

Db  2341 CGTCTCTACCTTGACAGACCTCCAGCCGTACATGCGACAGTTCGTGGCTCACCTGCAGGA 2400
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2341 CGTCTCTACCTTGACAGACCTCCAGCCGTACATGCGACAGTTCGTGGCTCACCTGCAGGA 2400

Db  2401 GACCAGCCCGCTGAGGGATGCCGTCGTCATCGAGCAGAGCTCCTCCCTGAATGAGGCCAG 2460
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2401 GACCAGCCCGCTGAGGGATGCCGTCGTCATCGAGCAGAGCTCCTCCCTGAATGAGGCCAG 2460

Db  2461 CAGTGGCCTCTTCGACGTCTTCCTACGCTTCATGTGCCACCACGCCGTGCGCATCAGGGG 2520
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2461 CAGTGGCCTCTTCGACGTCTTCCTACGCTTCATGTGCCACCACGCCGTGCGCATCAGGGG 2520

Db  2521 CAAGTCCTACGTCCAGTGCCAGGGGATCCCGCAGGGCTCCATCCTCTCCACGCTGCTCTG 2580
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2521 CAAGTCCTACGTCCAGTGCCAGGGGATCCCGCAGGGCTCCATCCTCTCCACGCTGCTCTG 2580

Db  2581 CAGCCTGTGCTACGGCGACATGGAGAACAAGCTGTTTGCGGGGATTCGGCGGGACGGGCT 2640
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2581 CAGCCTGTGCTACGGCGACATGGAGAACAAGCTGTTTGCGGGGATTCGGCGGGACGGGCT 2640

Db  2641 GCTCCTGCGTTTGGTGGATGATTTCTTGTTGGTGACACCTCACCTCACCCACGCGAAAAC 2700
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2641 GCTCCTGCGTTTGGTGGATGATTTCTTGTTGGTGACACCTCACCTCACCCACGCGAAAAC 2700

Db  2701 CTTCCTCAGGACCCTGGTCCGAGGTGTCCCTGAGTATGGCTGCGTGGTGAACTTGCGGAA 2760
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2701 CTTCCTCAGGACCCTGGTCCGAGGTGTCCCTGAGTATGGCTGCGTGGTGAACTTGCGGAA 2760

Db  2761 GACAGTGGTGAACTTCCCTGTAGAAGACGAGGCCCTGGGTGGCACGGCTTTTGTTCAGAT 2820
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2761 GACAGTGGTGAACTTCCCTGTAGAAGACGAGGCCCTGGGTGGCACGGCTTTTGTTCAGAT 2820

Db  2821 GCCGGCCCACGGCCTATTCCCCTGGTGCGGCCTGCTGCTGGATACCCGGACCCTGGAGGT 2880
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2821 GCCGGCCCACGGCCTATTCCCCTGGTGCGGCCTGCTGCTGGATACCCGGACCCTGGAGGT 2880

Db  2881 GCAGAGCGACTACTCCAGCTATGCCCGGACCTCCATCAGAGCCAGTCTCACCTTCAACCG 2940
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2881 GCAGAGCGACTACTCCAGCTATGCCCGGACCTCCATCAGAGCCAGTCTCACCTTCAACCG 2940

Db  2941 CGGCTTCAAGGCTGGGAGGAACATGCGTCGCAAACTCTTTGGGGTCTTGCGGCTGAAGTG 3000
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2941 CGGCTTCAAGGCTGGGAGGAACATGCGTCGCAAACTCTTTGGGGTCTTGCGGCTGAAGTG 3000

Db  3001 TCACAGCCTGTTTCTGGATTTGCAGGTGAACAGCCTCCAGACGGTGTGCACCAACATCTA 3060
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3001 TCACAGCCTGTTTCTGGATTTGCAGGTGAACAGCCTCCAGACGGTGTGCACCAACATCTA 3060

Db  3061 CAAGATCCTCCTGCTGCAGGCGTACAGGTTTCACGCATGTGTGCTGCAGCTCCCATTTCA 3120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3061 CAAGATCCTCCTGCTGCAGGCGTACAGGTTTCACGCATGTGTGCTGCAGCTCCCATTTCA 3120

Db  3121 TCAGCAAGTTTGGAAGAACCCCACATTTTTCCTGCGCGTCATCTCTGACACGGCCTCCCT 3180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3121 TCAGCAAGTTTGGAAGAACCCCACATTTTTCCTGCGCGTCATCTCTGACACGGCCTCCCT 3180

Db  3181 CTGCTACTCCATCCTGAAAGCCAAGAACGCAGGGATGTCGCTGGGGGCCAAGGGCGCCGC 3240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3181 CTGCTACTCCATCCTGAAAGCCAAGAACGCAGGGATGTCGCTGGGGGCCAAGGGCGCCGC 3240

Db  3241 CGGCCCTCTGCCCTCCGAGGCCGTGCAGTGGCTGTGCCACCAAGCATTCCTGCTCAAGCT 3300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3241 CGGCCCTCTGCCCTCCGAGGCCGTGCAGTGGCTGTGCCACCAAGCATTCCTGCTCAAGCT 3300

Db  3301 GACTCGACACCGTGTCACCTACGTGCCACTCCTGGGGTCACTCAGGACAGCCCAGACGCA 3360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3301 GACTCGACACCGTGTCACCTACGTGCCACTCCTGGGGTCACTCAGGACAGCCCAGACGCA 3360

Db  3361 GCTGAGTCGGAAGCTCCCGGGGGACGACGCTGACTGCCCTGGAGGCCGCAGCCAACCCGGC 3420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3361 GCTGAGTCGGAAGCTCCCGGGGGACGACGCTGACTGCCCTGGAGGCCGCAGCCAACCCGGC 3420

Db  3421 ACTGCCCTCAGACTTCAAGACCATCCTGGACTGATGGCCACCCGCCCACAGCCAGGCCGA 3480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3421 ACTGCCCTCAGACTTCAAGACCATCCTGGACTGATGGCCACCCGCCCACAGCCAGGCCGA 3480

Db  3481 GAGCAGACACCAGCAGCCCTGTCACGCCGGGCTCTACGTCCCAGGGAGGGAGGGGCGGCC 3540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3481 GAGCAGACACCAGCAGCCCTGTCACGCCGGGCTCTACGTCCCAGGGAGGGAGGGGCGGCC 3540

Db  3541 CACACCCAGGCCCGCACCGCTGGGAGTCTGAGGCCTGAGTGAGTGTTTGGCCGAGGCCTG 3600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3541 CACACCCAGGCCCGCACCGCTGGGAGTCTGAGGCCTGAGTGAGTGTTTGGCCGAGGCCTG 3600

Db  3601 CATGTCCGGCTGAAGGCTGAGTGTCCGGCTGAGGCCTGAGCGAGTGTCCAGCCAAGGGCT 3660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3601 CATGTCCGGCTGAAGGCTGAGTGTCCGGCTGAGGCCTGAGCGAGTGTCCAGCCAAGGGCT 3660

Db  3661 GAGTGTCCAGCACACCTGCCGTCTTCACTTCCCCACAGGCTGGCGCTCGGCTCCACCCCA 3720
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3661 GAGTGTCCAGCACACCTGCCGTCTTCACTTCCCCACAGGCTGGCGCTCGGCTCCACCCCA 3720

Db  3721 GGGCCAGCTTTTCCTCACCAGGAGCCCGGCTTCCACTCCCCACATAGGAATAGTCCATCC 3780
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3721 GGGCCAGCTTTTCCTCACCAGGAGCCCGGCTTCCACTCCCCACATAGGAATAGTCCATCC 3780

Db  3781 CCAGATTCGCCATTGTTCACCCCTCGCCCTGCCCTCCTTTGCCTTCCACCCCCACCATCC 3840
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3781 CCAGATTCGCCATTGTTCACCCCTCGCCCTGCCCTCCTTTGCCTTCCACCCCCACCATCC 3840

Db  3841 AGGTGGAGACCCTGAGAAGGACCCTGGGAGCTCTGGGAATTTGGAGTGACCAAAGGTGTG 3900
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3841 AGGTGGAGACCCTGAGAAGGACCCTGGGAGCTCTGGGAATTTGGAGTGACCAAAGGTGTG 3900

Db  3901 CCCTGTACACAGGCGAGGACCCTGCACCTGGATGGGGGTCCCTGTGGGTCAAATTGGGGG 3960
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3901 CCCTGTACACAGGCGAGGACCCTGCACCTGGATGGGGGTCCCTGTGGGTCAAATTGGGGG 3960

Db  3961 GAGGTGCTGTGGGAGTAAAATACTGAATATATGAGTTTTTCAGTTTTGAAAAAAA 4015
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy   3961 GAGGTGCTGTGGGAGTAAAATACTGAATATATGAGTTTTTCAGTTTTGAAAAAAA 4015


RESULT    2
LOCUS       AF018167     4027 bp    mRNA            PRI       29-AUG-1997
DEFINITION  Homo sapiens telomerase catalytic subunit (hEST2) mRNA, complete
            cds.
ACCESSION   AF018167
NID         g2347128
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 4027)
  AUTHORS   Meyerson,M., Counter,C.M., Eaton,E.N., Ellisen,L.W., Steiner,P.,
            Caddle,S.D., Ziaugra,L., Beijersbergen,R.L., Davidoff,M.J., Liu,Q.,
            Bacchetti,S., Haber,D.A. and Weinberg,R.A.
  TITLE     hEST2, the putative human telomerase catalytic subunit gene, is
            up-regulated in tumor cells and during immortalization
  JOURNAL   Cell 90 (4), 785-795 (1997)
  MEDLINE   97433088
REFERENCE   2  (bases 1 to 4027)
  AUTHORS   Meyerson,M., Counter,C.M., Eaton,E.N., Ellisen,L.W., Steiner,P.,
            Caddle,S.D., Ziaugra,L., Beijersbergen,R.L., Davidoff,M.J., Liu,Q.,
            Bacchetti,S., Haber,D.A. and Weinberg,R.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-AUG-1997) Whitehead Institute for Biomedical
            Research, Massachusetts Institute of Technology, Nine Cambridge
            Center, Cambridge, MA 02142, USA
FEATURES             Location/Qualifiers
     source          1. .4027
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5p15.33"
     gene            1. .4027
                     /note="human Ever Shorter Telomeres 2"
                     /gene="hEST2"
     CDS             57. .3455
                     /gene="hEST2"
                     /note="member of reverse transcriptase class"
                     /codon_start=1
                     /product="telomerase catalytic subunit"
                     /db_xref="PID:g2347129"
                     /translation="MPRAPRCRAVRSLLRSHYREVLPLATFVRRLGPQGWRLVQRGDP
                     AAFRALVAQCLVCVPWDARPPPAAPSFRQVSCLKELVARVLQRLCERGAKNVLAFGFA
                     LLDGARGGPPEAFTTSVRSYLPNTVTDALRGSGAWGLLLRRVGDDVLVHLLARCALFV
                     LVAPSCAYQVCGPPLYQLGAATQARPPPHASGPRRRLGCERAWNHSVREAGVPLGLPA
                     PGARRRGGSASRSLPLPKRPRRGAAPEPERTPVGQGSWAHPGRTRGPSDRGFCVVSPA
                     RPAEEATSLEGALSGTRHSHPSVGRQHHAGPPSTSRPPRPWDTPCPPVYAETKHFLYS
                     SGDKEQLRPSFLLSSLRPSLTGARRLVETIFLGSRPWMPGTPRRLPRLPQRYWQMRPL
                     FLELLGNHAQCPYGVLLKTHCPLRAAVTPAAGVCAREKPQGSVAAPEEEDTDPRRLVQ
                     LLRQHSSPWQVYGFVRACLRRLVPPGLWGSRHNERRFLRNTKKFISLGKHAKLSLQEL
                     TWKMSVRGCAWLRRSPGVGCVPAAEHRLREEILAKFLHWLMSVYVVELLRSFFYVTET
                     TFQKNRLFFYRKSVWSKLQSIGIRQHLKRVQLRELSEAEVRQHREARPALLTSRLRFI
                     PKPDGLRPIVNMDYVVGARTFRREKRAERLTSRVKALFSVLNYERARRPGLLGASVLG
                     LDDIHRAWRTFVLRVRAQDPPPELYFVKVDVTGAYDTIPQDRLTEVIASIIKPQNTYC
                     VRRYAVVQKAAHGHVRKAFKSHVSTLTDLQPYMRQFVAHLQETSPLRDAVVIEQSSSL
                     NEASSGLFDVFLRFMCHHAVRIRGKSYVQCQGIPQGSILSTLLCSLCYGDMENKLFAG
                     IRRDGLLLRLVDDFLLVTPHLTHAKTFLRTLVRGVPEYGCVVNLRKTVVNFPVEDEAL
                     GGTAFVQMPAHGLFPWCGLLLDTRTLEVQSDYSSYARTSIRASLTFNRGFKAGRNMRR
                     KLFGVLRLKCHSLFLDLQVNSLQTVCTNIYKILLLQAYRFHACVLQLPFHQQVWKNPT
                     FFLRVISDTASLCYSILKAKNAGMSLGAKGAAGPLPSEAVQWLCHQAFLLKLTRHRVT
                     YVPLLGSLRTAQTQLSRKLPGTTLTALEAAANPALPSDFKTILD"
BASE COUNT      674 a   1361 c   1277 g    715 t
ORIGIN

  Query Match           99.7%;  Score 4001;  DB 21;  Length 4027;
  Best Local Similarity 100.0%;  Pred. No. 0.00e+00;
  Matches  4003;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;
```

```
Db     12 GTCCTGCTGCGCACGTGGGAAGCCCTGGCCCCGGCCACCCCCGCGATGCCGCGCGCTCCC 71
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     11 GTCCTGCTGCGCACGTGGGAAGCCCTGGCCCCGGCCACCCCCGCGATGCCGCGCGCTCCC 70

Db     72 CGCTGCCGAGCCGTGCGCTCCCTGCTGCGCAGCCACTACCGCGAGGTGCTGCCGCTGGCC 131
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     71 CGCTGCCGAGCCGTGCGCTCCCTGCTGCGCAGCCACTACCGCGAGGTGCTGCCGCTGGCC 130

Db    132 ACGTTCGTGCGGCGCCTGGGGCCCCAGGGCTGGCGGCTGGTGCAGCGCGGGGACCCGGCG 191
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    131 ACGTTCGTGCGGCGCCTGGGGCCCCAGGGCTGGCGGCTGGTGCAGCGCGGGGACCCGGCG 190

Db    192 GCTTTCCGCGCGCTGGTGGCCCAGTGCCTGGTGTGCGTGCCCTGGGACGCACGGCCGCCC 251
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    191 GCTTTCCGCGCGCTGGTGGCCCAGTGCCTGGTGTGCGTGCCCTGGGACGCACGGCCGCCC 250

Db    252 CCCGCCGCCCCCTCCTTCCGCCAGGTGTCCTGCCTGAAGGAGCTGGTGGCCCGAGTGCTG 311
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    251 CCCGCCGCCCCCTCCTTCCGCCAGGTGTCCTGCCTGAAGGAGCTGGTGGCCCGAGTGCTG 310

Db    312 CAGAGGCTGTGCGAGCGCGGCGCGAAGAACGTGCTGGCCTTCGGCTTCGCGCTGCTGGAC 371
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    311 CAGAGGCTGTGCGAGCGCGGCGCGAAGAACGTGCTGGCCTTCGGCTTCGCGCTGCTGGAC 370

Db    372 GGGGCCCGCGGGGGCCCCCCCGAGGCCTTCACCACCAGCGTGCGCAGCTACCTGCCCAAC 431
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    371 GGGGCCCGCGGGGGCCCCCCCGAGGCCTTCACCACCAGCGTGCGCAGCTACCTGCCCAAC 430

Db    432 ACGGTGACCGACGCACTGCGGGGGAGCGGGGCGTGGGGGCTGCTGTTGCGCCGCGTGGGC 491
        |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Qy    431 ACGGTGACCGACGCACTGCGGGGGAGCGGGGCGTGGGGGCTGCTGCTGCGCCGCGTGGGC 490

Db    492 GACGACGTGCTGGTTCACCTGCTGGCACGCTGCGCGCTCTTTGTGCTGGTGGCTCCCAGC 551
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    491 GACGACGTGCTGGTTCACCTGCTGGCACGCTGCGCGCTCTTTGTGCTGGTGGCTCCCAGC 550

Db    552 TGCGCCTACCAGGTGTGCGGGCCGCCGCTGTACCAGCTCGGCGCTGCCACTCAGGCCCGG 611
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    551 TGCGCCTACCAGGTGTGCGGGCCGCCGCTGTACCAGCTCGGCGCTGCCACTCAGGCCCGG 610

Db    612 CCCCCGCCACACGCTAGTGGACCCCGAAGGCGTCTGGGATGCGAACGGGCCTGGAACCAT 671
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    611 CCCCCGCCACACGCTAGTGGACCCCGAAGGCGTCTGGGATGCGAACGGGCCTGGAACCAT 670

Db    672 AGCGTCAGGGAGGCCGGGGTCCCCCTGGGCCTGCCAGCCCCGGGTGCGAGGAGGCGCGGG 731
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    671 AGCGTCAGGGAGGCCGGGGTCCCCCTGGGCCTGCCAGCCCCGGGTGCGAGGAGGCGCGGG 730

Db    732 GGCAGTGCCAGCCGAAGTCTGCCGTTGCCCAAGAGGCCCAGGCGTGGCGCTGCCCCTGAG 791
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    731 GGCAGTGCCAGCCGAAGTCTGCCGTTGCCCAAGAGGCCCAGGCGTGGCGCTGCCCCTGAG 790

Db    792 CCGGAGCGGACGCCCGTTGGGCAGGGGTCCTGGGCCCACCCGGGCAGGACGCGTGGACCG 851
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    791 CCGGAGCGGACGCCCGTTGGGCAGGGGTCCTGGGCCCACCCGGGCAGGACGCGTGGACCG 850

Db    852 AGTGACCGTGGTTTCTGTGTGGTGTCACCTGCCAGACCCGCCGAAGAAGCCACCTCTTTG 911
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    851 AGTGACCGTGGTTTCTGTGTGGTGTCACCTGCCAGACCCGCCGAAGAAGCCACCTCTTTG 910

Db    912 GAGGGTGCGCTCTCTGGCACGCGCCACTCCCACCCATCCGTGGGCCGCCAGCACCACGCG 971
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    911 GAGGGTGCGCTCTCTGGCACGCGCCACTCCCACCCATCCGTGGGCCGCCAGCACCACGCG 970

Db    972 GGCCCCCCATCCACATCGCGGCCACCACGTCCCTGGGACACGCCTTGTCCCCCGGTGTAC 1031
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    971 GGCCCCCCATCCACATCGCGGCCACCACGTCCCTGGGACACGCCTTGTCCCCCGGTGTAC 1030

Db   1032 GCCGAGACCAAGCACTTCCTCTACTCCTCAGGCGACAAGGAGCAGCTGCGGCCCTCCTTC 1091
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1031 GCCGAGACCAAGCACTTCCTCTACTCCTCAGGCGACAAGGAGCAGCTGCGGCCCTCCTTC 1090

Db   1092 CTACTCAGCTCTCTGAGGCCCAGCCTGACTGGCGCTCGGAGGCTCGTGGAGACCATCTTT 1151
```

```
Qy   1091 CTACTCAGCTCTCTGAGGCCCAGCCTGACTGGCGCTCGGAGGCTCGTGGAGACCATCTTT 1150
Db   1152 CTGGGTTCCAGGCCCTGGATGCCAGGGACTCCCCGCAGGTTGCCCCGCCTGCCCCAGCGC 1211
Qy   1151 CTGGGTTCCAGGCCCTGGATGCCAGGGACTCCCCGCAGGTTGCCCCGCCTGCCCCAGCGC 1210
Db   1212 TACTGGCAAATGCGGCCCCTGTTTCTGGAGCTGCTTGGGAACCACGCGCAGTGCCCCTAC 1271
Qy   1211 TACTGGCAAATGCGGCCCCTGTTTCTGGAGCTGCTTGGGAACCACGCGCAGTGCCCCTAC 1270
Db   1272 GGGGTGCTCCTCAAGACGCACTGCCCGCTGCGAGCTGCGGTCACCCCAGCAGCCGGTGTC 1331
Qy   1271 GGGGTGCTCCTCAAGACGCACTGCCCGCTGCGAGCTGCGGTCACCCCAGCAGCCGGTGTC 1330
Db   1332 TGTGCCCGGGAGAAGCCCCAGGGCTCTGTGGCGGCCCCCGAGGAGGAGGACACAGACCCC 1391
Qy   1331 TGTGCCCGGGAGAAGCCCCAGGGCTCTGTGGCGGCCCCCGAGGAGGAGGACACAGACCCC 1390
Db   1392 CGTCGCCTGGTGCAGCTGCTCCGCCAGCACAGCAGCCCCTGGCAGGTGTACGGCTTCGTG 1451
Qy   1391 CGTCGCCTGGTGCAGCTGCTCCGCCAGCACAGCAGCCCCTGGCAGGTGTACGGCTTCGTG 1450
Db   1452 CGGGCCTGCCTGCGCCGGCTGGTGCCCCCAGGCCTCTGGGGCTCCAGGCACAACGAACGC 1511
Qy   1451 CGGGCCTGCCTGCGCCGGCTGGTGCCCCCAGGCCTCTGGGGCTCCAGGCACAACGAACGC 1510
Db   1512 CGCTTCCTCAGGAACACCAAGAAGTTCATCTCCCTGGGGAAGCATGCCAAGCTCTCGCTG 1571
Qy   1511 CGCTTCCTCAGGAACACCAAGAAGTTCATCTCCCTGGGGAAGCATGCCAAGCTCTCGCTG 1570
Db   1572 CAGGAGCTGACGTGGAAGATGAGCGTGCGGGGCTGCGCTTGGCTGCGCAGGAGCCCAGGG 1631
Qy   1571 CAGGAGCTGACGTGGAAGATGAGCGTGCGGGACTGCGCTTGGCTGCGCAGGAGCCCAGGG 1630
Db   1632 GTTGGCTGTGTTCCGGCCGCAGAGCACCGTCTGCGTGAGGAGATCCTGGCCAAGTTCCTG 1691
Qy   1631 GTTGGCTGTGTTCCGGCCGCAGAGCACCGTCTGCGTGAGGAGATCCTGGCCAAGTTCCTG 1690
Db   1692 CACTGGCTGATGAGTGTGTACGTCGTCGAGCTGCTCAGGTCTTTCTTTTATGTCACGGAG 1751
Qy   1691 CACTGGCTGATGAGTGTGTACGTCGTCGAGCTGCTCAGGTCTTTCTTTTATGTCACGGAG 1750
Db   1752 ACCACGTTTCAAAAGAACAGGCTCTTTTTCTACCGGAAGAGTGTCTGGAGCAAGTTGCAA 1811
Qy   1751 ACCACGTTTCAAAAGAACAGGCTCTTTTTCTACCGGAAGAGTGTCTGGAGCAAGTTGCAA 1810
Db   1812 AGCATTGGAATCAGACAGCACTTGAAGAGGGTGCAGCTGCGGGAGCTGTCGGAAGCAGAG 1871
Qy   1811 AGCATTGGAATCAGACAGCACTTGAAGAGGGTGCAGCTGCGGGAGCTGTCGGAAGCAGAG 1870
Db   1872 GTCAGGCAGCATCGGGAAGCCAGGCCCGCCCTGCTGACGTCCAGACTCCGCTTCATCCCC 1931
Qy   1871 GTCAGGCAGCATCGGGAAGCCAGGCCCGCCCTGCTGACGTCCAGACTCCGCTTCATCCCC 1930
Db   1932 AAGCCTGACGGGCTGCGGCCGATTGTGAACATGGACTACGTCGTGGGAGCCAGAACGTTC 1991
Qy   1931 AAGCCTGACGGGCTGCGGCCGATTGTGAACATGGACTACGTCGTGGGAGCCAGAACGTTC 1990
Db   1992 CGCAGAGAAAAGAGGGCCGAGCGTCTCACCTCGAGGGTGAAGGCACTGTTCAGCGTGCTC 2051
Qy   1991 CGCAGAGAAAAGAGGGCCGAGCGTCTCACCTCGAGGGTGAAGGCACTGTTCAGCGTGCTC 2050
Db   2052 AACTACGAGCGGGCGCGGCGCCCCGGCCTCCTGGGCGCCTCTGTGCTGGGCCTGGACGAT 2111
Qy   2051 AACTACGAGCGGGCGCGGCGCCCCGGCCTCCTGGGCGCCTCTGTGCTGGGCCTGGACGAT 2110
Db   2112 ATCCACAGGGCCTGGCGCACCTTCGTGCTGCGTGTGCGGGCCCAGGACCCGCCGCCTGAG 2171
Qy   2111 ATCCACAGGGCCTGGCGCACCTTCGTGCTGCGTGTGCGGGCCCAGGACCCGCCGCCTGAG 2170
Db   2172 CTGTACTTTGTCAAGGTGGATGTGACGGGCGCGTACGACACCATCCCCCAGGACAGGCTC 2231
Qy   2171 CTGTACTTTGTCAAGGTGGATGTGACGGGCGCGTACGACACCATCCCCCAGGACAGGCTC 2230
Db   2232 ACGGAGGTCATCGCCAGCATCATCAAACCCCAGAACACGTACTGCGTGCGTCGGTATGCC 2291
Qy   2231 ACGGAGGTCATCGCCAGCATCATCAAACCCCAGAACACGTACTGCGTGCGTCGGTATGCC 2290
Db   2292 GTGGTCCAGAAGGCCGCCCATGGGCACGTCCGCAAGGCCTTCAAGAGCCACGTCTCTACC 2351
Qy   2291 GTGGTCCAGAAGGCCGCCCATGGGCACGTCCGCAAGGCCTTCAAGAGCCACGTCTCTACC 2350
Db   2352 TTGACAGACCTCCAGCCGTACATGCGACAGTTCGTGGCTCACCTGCAGGAGACCAGCCCG 2411
Qy   2351 TTGACAGACCTCCAGCCGTACATGCGACAGTTCGTGGCTCACCTGCAGGAGACCAGCCCG 2410
Db   2412 CTGAGGGATGCCGTCGTCATCGAGCAGAGCTCCTCCCTGAATGAGGCCAGCAGTGGCCTC 2471
Qy   2411 CTGAGGGATGCCGTCGTCATCGAGCAGAGCTCCTCCCTGAATGAGGCCAGCAGTGGCCTC 2470
Db   2472 TTCGACGTCTTCCTACGCTTCATGTGCCACCACGCCGTGCGCATCAGGGGCAAGTCCTAC 2531
Qy   2471 TTCGACGTCTTCCTACGCTTCATGTGCCACCACGCCGTGCGCATCAGGGGCAAGTCCTAC 2530
Db   2532 GTCCAGTGCCAGGGGATCCCGCAGGGCTCCATCCTCTCCACGCTGCTCTGCAGCCTGTGC 2591
Qy   2531 GTCCAGTGCCAGGGGATCCCGCAGGGCTCCATCCTCTCCACGCTGCTCTGCAGCCTGTGC 2590
Db   2592 TACGGCGACATGGAGAACAAGCTGTTTGCGGGGATTCGGCGGGACGGGCTGCTCCTGCGT 2651
Qy   2591 TACGGCGACATGGAGAACAAGCTGTTTGCGGGGATTCGGCGGGACGGGCTGCTCCTGCGT 2650
Db   2652 TTGGTGGATGATTTCTTGTTGGTGACACCTCACCTCACCCACGCGAAAACCTTCCTCAGG 2711
Qy   2651 TTGGTGGATGATTTCTTGTTGGTGACACCTCACCTCACCCACGCGAAAACCTTCCTCAGG 2710
Db   2712 ACCCTGGTCCGAGGTGTCCCTGAGTATGGCTGCGTGGTGAACTTGCGGAAGACAGTGGTG 2771
Qy   2711 ACCCTGGTCCGAGGTGTCCCTGAGTATGGCTGCGTGGTGAACTTGCGGAAGACAGTGGTG 2770
Db   2772 AACTTCCCTGTAGAAGACGAGGCCCTGGGTGGCACGGCTTTTGTTCAGATGCCGGCCCAC 2831
Qy   2771 AACTTCCCTGTAGAAGACGAGGCCCTGGGTGGCACGGCTTTTGTTCAGATGCCGGCCCAC 2830
Db   2832 GGCCTATTCCCCTGGTGCGGCCTGCTGCTGGATACCCGGACCCTGGAGGTGCAGAGCGAC 2891
Qy   2831 GGCCTATTCCCCTGGTGCGGCCTGCTGCTGGATACCCGGACCCTGGAGGTGCAGAGCGAC 2890
Db   2892 TACTCCAGCTATGCCCGGACCTCCATCAGAGCCAGTCTCACCTTCAACCGCGGCTTCAAG 2951
Qy   2891 TACTCCAGCTATGCCCGGACCTCCATCAGAGCCAGTCTCACCTTCAACCGCGGCTTCAAG 2950
Db   2952 GCTGGGAGGAACATGCGTCGCAAACTCTTTGGGGTCTTGCGGCTGAAGTGTCACAGCCTG 3011
Qy   2951 GCTGGGAGGAACATGCGTCGCAAACTCTTTGGGGTCTTGCGGCTGAAGTGTCACAGCCTG 3010
Db   3012 TTTCTGGATTTGCAGGTGAACAGCCTCCAGACGGTGTGCACCAACATCTACAAGATCCTC 3071
Qy   3011 TTTCTGGATTTGCAGGTGAACAGCCTCCAGACGGTGTGCACCAACATCTACAAGATCCTC 3070
Db   3072 CTGCTGCAGGCGTACAGGTTTCACGCATGTGTGCTGCAGCTCCCATTTCATCAGCAAGTT 3131
Qy   3071 CTGCTGCAGGCGTACAGGTTTCACGCATGTGTGCTGCAGCTCCCATTTCATCAGCAAGTT 3130
Db   3132 TGGAAGAACCCCACATTTTTCCTGCGCGTCATCTCTGACACGGCCTCCCTCTGCTACTCC 3191
Qy   3131 TGGAAGAACCCCACATTTTTCCTGCGCGTCATCTCTGACACGGCCTCCCTCTGCTACTCC 3190
Db   3192 ATCCTGAAAGCCAAGAACGCAGGGATGTCGCTGGGGGCCAAGGGCGCCGCCGGCCCTCTG 3251
Qy   3191 ATCCTGAAAGCCAAGAACGCAGGGATGTCGCTGGGGGCCAAGGGCGCCGCCGGCCCTCTG 3250
Db   3252 CCCTCCGAGGCCGTGCAGTGGCTGTGCCACCAAGCATTCCTGCTCAAGCTGACTCGACAC 3311
Qy   3251 CCCTCCGAGGCCGTGCAGTGGCTGTGCCACCAAGCATTCCTGCTCAAGCTGACTCGACAC 3310
```

```
Db   3312 CGTGTCACCTACGTGCCACTCCTGGGGTCACTCAGGACAGCCCAGACGCAGCTGAGTCGG 3371
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3311 CGTGTCACCTACGTGCCACTCCTGGGGTCACTCAGGACAGCCCAGACGCAGCTGAGTCGG 3370

Db   3372 AAGCTCCCGGGGACGACGCTGACTGCCCTGGAGGCCGCAGCCAACCCGGCACTGCCCTCA 3431
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3371 AAGCTCCCGGGGACGACGCTGACTGCCCTGGAGGCCGCAGCCAACCCGGCACTGCCCTCA 3430

Db   3432 GACTTCAAGACCATCCTGGACTGATGGCCACCCGCCCACAGCCAGGCCGAGAGCAGACAC 3491
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3431 GACTTCAAGACCATCCTGGACTGATGGCCACCCGCCCACAGCCAGGCCGAGAGCAGACAC 3490

Db   3492 CAGCAGCCCTGTCACGCCGGGCTCTACGTCCCAGGGAGGGAGGGGCGGCCCACACCCAGG 3551
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3491 CAGCAGCCCTGTCACGCCGGGCTCTACGTCCCAGGGAGGGAGGGGCGGCCCACACCCAGG 3550

Db   3552 CCCGCACCGCTGGGAGTCTGAGGCCTGAGTGAGTGTTTGGCCGAGGCCTGCATGTCCGGC 3611
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3551 CCCGCACCGCTGGGAGTCTGAGGCCTGAGTGAGTGTTTGGCCGAGGCCTGCATGTCCGGC 3610

Db   3612 TGAAGGCTGAGTGTCCGGCTGAGGCCTGAGCGAGTGTCCAGCCAAGGGCTGAGTGTCCAG 3671
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3611 TGAAGGCTGAGTGTCCGGCTGAGGCCTGAGCGAGTGTCCAGCCAAGGGCTGAGTGTCCAG 3670

Db   3672 CACACCTGCCGTCTTCACTTCCCCACAGGCTGGCGCTCGGCTCCACCCCAGGGCCAGCTT 3731
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3671 CACACCTGCCGTCTTCACTTCCCCACAGGCTGGCGCTCGGCTCCACCCCAGGGCCAGCTT 3730

Db   3732 TTCCTCACCAGGAGCCCGGCTTCCACTCCCCACATAGGAATAGTCCATCCCCAGATTCGC 3791
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3731 TTCCTCACCAGGAGCCCGGCTTCCACTCCCCACATAGGAATAGTCCATCCCCAGATTCGC 3790

Db   3792 CATTGTTCACCCCTCGCCCTGCCCTCCTTTGCCTTCCACCCCCACCATCCAGGTGGAGAC 3851
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3791 CATTGTTCACCCCTCGCCCTGCCCTCCTTTGCCTTCCACCCCCACCATCCAGGTGGAGAC 3850

Db   3852 CCTGAGAAGGACCCTGGGAGCTCTGGGAATTTGGAGTGACCAAAGGTGTGCCCTGTACAC 3911
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3851 CCTGAGAAGGACCCTGGGAGCTCTGGGAATTTGGAGTGACCAAAGGTGTGCCCTGTACAC 3910

Db   3912 AGGCGAGGACCCTGCACCTGGATGGGGGTCCCTGTGGGTCAAATTGGGGGGAGGTGCTGT 3971
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3911 AGGCGAGGACCCTGCACCTGGATGGGGGTCCCTGTGGGTCAAATTGGGGGGAGGTGCTGT 3970

Db   3972 GGGAGTAAAATACTGAATATATGAGTTTTTCAGTTTTGAAAAAAA 4016
          |||||||||||||||||||||||||||||||||||||||||||
Qy   3971 GGGAGTAAAATACTGAATATATGAGTTTTTCAGTTTTGAAAAAAA 4015
```

RESULT 3
LOCUS AF043739 8960 bp DNA SYN 05-APR-1998
DEFINITION Synthetic construct human telomerase catalytic subunit (hTERT/hEST2) mRNA and 3' Influenza A hemagglutinin tag, complete sequence.
ACCESSION AF043739
NID g3023054
KEYWORDS .
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequence.
REFERENCE 1 (bases 1 to 8960)
AUTHORS Counter,C.M., Meyerson,M., Eaton,E.N., Ellisen,L.W., Caddle,S.D., Haber,D.A. and Weinberg,R.A.
TITLE Telomerase activity is restored in human cells by ectopic expression of hTERT (hEST2), the catalytic subunit of telomerase
JOURNAL Oncogene 16 (9), 1217-1222 (1998)
MEDLINE 98187713
REFERENCE 2 (bases 1 to 8960)
AUTHORS Counter,C.M., Meyerson,M., Eaton,E.N., Ellisen,L.W., Caddle,S.D., Haber,D. and Weinberg,R.A.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-1998) Whitehead Institute, 9 Cambridge Center, Cambridge, MA 02142, USA

too recent

```
FEATURES             Location/Qualifiers
     source          1. .8960
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
     source          order(1. .1101,4606. .8960)
                     /note="mammalian expression vector; see GenBank Accession
                     Number U47120"
                     /organism="Cloning vector pCI-neo"
                     /db_xref="taxon:45772"
     source          1102. .4542
                     /note="telomerase catalytic subunit hTERT/hEST2 mRNA"
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5p15.33"
     gene            1102. .4542
                     /gene="hTERT/hEST2"
     CDS             1147. .4605
                     /note="telomerase/reverse transcriptase"
                     /codon_start=1
                     /transl_table=11
                     /product="telomerase catalytic subunit"
                     /db_xref="PID:g3023055"
                     /translation="MPRAPRCRAVRSLLRSHYREVLPLATFVRRLGPQGWRLVQRGDP
                     AAFRALVAQCLVCVPWDARPPPAAPSFRQVSCLKELVARVLQRLCERGAKNVLAFGFA
                     LLDGARGGPPEAFTTSVRSYLPNTVTDALRGSGAWGLLLRRVGDDVLVHLLARCALFV
                     LVAPSCAYQVCGPPLYQLGAATQARPPPHASGPRRRLGCERAWNHSVREAGVPLGLPA
                     PGARRRGGSASRSLPLPKRPRRGAAPEPERTPVGQGSWAHPGRTRGPSDRGFCVVSPA
                     RPAEEATSLEGALSGTRHSHPSVGRQHHAGPPSTSRPPRPWDTPCPPVYAETKHFLYS
                     SGDKEQLRPSFLLSSLRPSLTGARRLVETIFLGSRPWMPGTPRRLPRLPQRYWQMRPL
                     FLELLGNHAQCPYGVLLKTHCPLRAAVTPAAGVCAREKPQGSVAAPEEEDTDPRRLVQ
                     LLRQHSSPWQVYGFVRACLRRLVPPGLWGSRHNERRFLRNTKKFISLGKHAKLSLQEL
                     TWKMSVRGCAWLRRSPGVGCVPAAEHRLREEILAKFLHWLMSVYVVELLRSFFYVTET
                     TFQKNRLFFYRKSVWSKLQSIGIRQHLKRVQLRELSEAEVRQHREARPALLTSRLRFI
                     PKPDGLRPIVNMDYVVGARTFRREKRAERLTSRVKALFSVLNYERARRPGLLGASVLG
                     LDDIHRAWRTFVLRVRAQDPPPELYFVKVDVTGAYDTIPQDRLTEVIASIIKPQNTYC
                     VRRYAVVQKAAHGHVRKAFKSHVSTLTDLQPYMRQFVAHLQETSPLRDAVVIEQSSSL
                     NEASSGLFDVFLRFMCHHAVRIRGKSYVQCQGIPQGSILSTLLCSLCYGDMENKLFAG
                     IRRDGLLLRLVDDFLLVTPHLTHAKTFLRTLVRGVPEYGCVVNLRKTVVNFPVEDEAL
                     GGTAFVQMPAHGLFPWCGLLLDTRTLEVQSDYSSYARTSIRASLTFNRGFKAGRNMRR
                     KLFGVLRLKCHSLFLDLQVNSLQTVCTNIYKILLLQAYRFHACVLQLPFHQQVWKNPT
                     FFLRVISDTASLCYSILKAKNAGMSLGAKGAAGPLPSEAVQWLCHQAFLLKLTRHRVT
                     YVPLLGSLRTAQTQLSRKLPGTTLTALEAAANPALPSDFKTILDYPYDVPDYAGYPYD
                     VPDYAG"
     source          4543. .4605
                     /note="engineered based on the epitope for the
                     anti-influenza virus hemagglutinin monoclonal antibody
                     12CA5; similar to part of the Influenza A hemaglutinin
                     gene sequence in GenBank Accession Number M29257"
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
BASE COUNT 1886 a 2551 c 2484 g 2038 t 1 others
ORIGIN
```

Query Match 85.7%; Score 3439; DB 25; Length 8960;
Best Local Similarity 99.9%; Pred. No. 0.00e+00;
Matches 3441; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Db   1101 CGTCCTGCTGCGCACGTGGGAAGCCCTGGCCCCGGCCACCCCCGCGATGCCGCGCGCTCC 1160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     10 CGTCCTGCTGCGCACGTGGGAAGCCCTGGCCCCGGCCACCCCCGCGATGCCGCGCGCTCC 69

Db   1161 CCGCTGCCGAGCCGTGCGCTCCCTGCTGCGCAGCCACTACCGCGAGGTGCTGCCGCTGGC 1220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     70 CCGCTGCCGAGCCGTGCGCTCCCTGCTGCGCAGCCACTACCGCGAGGTGCTGCCGCTGGC 129

Db   1221 CACGTTCGTGCGGCGCCTGGGGCCCCAGGGCTGGCGGCTGGTGCAGCGCGGGGACCCGGC 1280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    130 CACGTTCGTGCGGCGCCTGGGGCCCCAGGGCTGGCGGCTGGTGCAGCGCGGGGACCCGGC 189

Db   1281 GGCTTTCCGCGCGCTGGTGGCCCAGTGCCTGGTGTGCGTGCCCTGGGACGCACGGCCGCC 1340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy    190 GGCTTTCCGCGCGCTGGTGGCCCAGTGCCTGGTGTGCGTGCCCTGGGACGCACGGCCGCC 249

Db   1341 CCCCGCCGCCCCCTCCTTCCGCCAGGTGTCCTGCCTGAAGGAGCTGGTGGCCCGAGTGCT 1400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    250 CCCCGCCGCCCCCTCCTTCCGCCAGGTGTCCTGCCTGAAGGAGCTGGTGGCCCGAGTGCT 309

Db   1401 GCAGAGGCTGTGCGAGCGCGGCGCGAAGAACGTGCTGGCCTTCGGCTTCGCGCTGCTGGA 1460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    310 GCAGAGGCTGTGCGAGCGCGGCGCGAAGAACGTGCTGGCCTTCGGCTTCGCGCTGCTGGA 369

Db   1461 CGGGGCCCGCGGGGGCCCCCCCGAGGCCTTCACCACCAGCGTGCGCAGCTACCTGCCCAA 1520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    370 CGGGGCCCGCGGGGGCCCCCCCGAGGCCTTCACCACCAGCGTGCGCAGCTACCTGCCCAA 429

Db   1521 CACGGTGACCGACGCACTGCGGGGGAGCGGGGCGTGGGGGCTGCTGTTGCGCCGCGTGGG 1580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    430 CACGGTGACCGACGCACTGCGGGGGAGCGGGGCGTGGGGGCTGCTGCTGCGCCGCGTGGG 489

Db   1581 CGACGACGTGCTGGTTCACCTGCTGGCACGCTGCGCGCTCTTTGTGCTGGTGGCTCCCAG 1640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    490 CGACGACGTGCTGGTTCACCTGCTGGCACGCTGCGCGCTCTTTGTGCTGGTGGCTCCCAG 549

Db   1641 CTGCGCCTACCAGGTGTGCGGGCCGCCGCTGTACCAGCTCGGCGCTGCCACTCAGGCCCG 1700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    550 CTGCGCCTACCAGGTGTGCGGGCCGCCGCTGTACCAGCTCGGCGCTGCCACTCAGGCCCG 609

Db   1701 GCCCCCGCCACACGCTAGTGGACCCCGAAGGCGTCTGGGATGCGAACGGGCCTGGAACCA 1760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    610 GCCCCCGCCACACGCTAGTGGACCCCGAAGGCGTCTGGGATGCGAACGGGCCTGGAACCA 669

Db   1761 TAGCGTCAGGGAGGCCGGGGTCCCCCTGGGCCTGCCAGCCCCGGGTGCGAGGAGGCGCGG 1820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    670 TAGCGTCAGGGAGGCCGGGGTCCCCCTGGGCCTGCCAGCCCCGGGTGCGAGGAGGCGCGG 729

Db   1821 GGGCAGTGCCAGCCGAAGTCTGCCGTTGCCCAAGAGGCCCAGGCGTGGCGCTGCCCCTGA 1880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    730 GGGCAGTGCCAGCCGAAGTCTGCCGTTGCCCAAGAGGCCCAGGCGTGGCGCTGCCCCTGA 789

Db   1881 GCCGGAGCGGACGCCCGTTGGGCAGGGGTCCTGGGCCCACCCGGGCAGGACGCGTGGACC 1940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    790 GCCGGAGCGGACGCCCGTTGGGCAGGGGTCCTGGGCCCACCCGGGCAGGACGCGTGGACC 849

Db   1941 GAGTGACCGTGGTTTCTGTGTGGTGTCACCTGCCAGACCCGCCGAAGAAGCCACCTCTTT 2000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    850 GAGTGACCGTGGTTTCTGTGTGGTGTCACCTGCCAGACCCGCCGAAGAAGCCACCTCTTT 909

Db   2001 GGAGGGTGCGCTCTCTGGCACGCGCCACTCCCACCCATCCGTGGGCCGCCAGCACCACGC 2060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    910 GGAGGGTGCGCTCTCTGGCACGCGCCACTCCCACCCATCCGTGGGCCGCCAGCACCACGC 969

Db   2061 GGGCCCCCCATCCACATCGCGGCCACCACGTCCCTGGGACACGCCTTGTCCCCCGGTGTA 2120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    970 GGGCCCCCCATCCACATCGCGGCCACCACGTCCCTGGGACACGCCTTGTCCCCCGGTGTA 1029

Db   2121 CGCCGAGACCAAGCACTTCCTCTACTCCTCAGGCGACAAGGAGCAGCTGCGGCCCTCCTT 2180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1030 CGCCGAGACCAAGCACTTCCTCTACTCCTCAGGCGACAAGGAGCAGCTGCGGCCCTCCTT 1089

Db   2181 CCTACTCAGCTCTCTGAGGCCCAGCCTGACTGGCGCTCGGAGGCTCGTGGAGACCATCTT 2240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1090 CCTACTCAGCTCTCTGAGGCCCAGCCTGACTGGCGCTCGGAGGCTCGTGGAGACCATCTT 1149

Db   2241 TCTGGGTTCCAGGCCCTGGATGCCAGGGACTCCCCGCAGGTTGCCCCGCCTGCCCCAGCG 2300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1150 TCTGGGTTCCAGGCCCTGGATGCCAGGGACTCCCCGCAGGTTGCCCCGCCTGCCCCAGCG 1209

Db   2301 CTACTGGCAAATGCGGCCCCTGTTTCTGGAGCTGCTTGGGAACCACGCGCAGTGCCCCTA 2360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1210 CTACTGGCAAATGCGGCCCCTGTTTCTGGAGCTGCTTGGGAACCACGCGCAGTGCCCCTA 1269

Db   2361 CGGGGTGCTCCTCAAGACGCACTGCCCGCTGCGAGCTGCGGTCACCCCAGCAGCCGGTGT 2420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1270 CGGGGTGCTCCTCAAGACGCACTGCCCGCTGCGAGCTGCGGTCACCCCAGCAGCCGGTGT 1329

Db   2421 CTGTGCCCGGGAGAAGCCCCAGGGCTCTGTGGCGGCCCCCGAGGAGGAGGACACAGACCC 2480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1330 CTGTGCCCGGGAGAAGCCCCAGGGCTCTGTGGCGGCCCCCGAGGAGGAGGACACAGACCC 1389

Db   2481 CCGTCGCCTGGTGCAGCTGCTCCGCCAGCACAGCAGCCCCTGGCAGGTGTACGGCTTCGT 2540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1390 CCGTCGCCTGGTGCAGCTGCTCCGCCAGCACAGCAGCCCCTGGCAGGTGTACGGCTTCGT 1449

Db   2541 GCGGGCCTGCCTGCGCCGGCTGGTGCCCCCAGGCCTCTGGGGCTCCAGGCACAACGAACG 2600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1450 GCGGGCCTGCCTGCGCCGGCTGGTGCCCCCAGGCCTCTGGGGCTCCAGGCACAACGAACG 1509

Db   2601 CCGCTTCCTCAGGAACACCAAGAAGTTCATCTCCCTGGGGAAGCATGCCAAGCTCTCGCT 2660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1510 CCGCTTCCTCAGGAACACCAAGAAGTTCATCTCCCTGGGGAAGCATGCCAAGCTCTCGCT 1569

Db   2661 GCAGGAGCTGACGTGGAAGATGAGCGTGCGGGGCTGCGCTTGGCTGCGCAGGAGCCCAGG 2720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1570 GCAGGAGCTGACGTGGAAGATGAGCGTGCGGGACTGCGCTTGGCTGCGCAGGAGCCCAGG 1629

Db   2721 GGTTGGCTGTGTTCCGGCCGCAGAGCACCGTCTGCGTGAGGAGATCCTGGCCAAGTTCCT 2780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1630 GGTTGGCTGTGTTCCGGCCGCAGAGCACCGTCTGCGTGAGGAGATCCTGGCCAAGTTCCT 1689

Db   2781 GCACTGGCTGATGAGTGTGTACGTCGTCGAGCTGCTCAGGTCTTTCTTTTATGTCACGGA 2840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1690 GCACTGGCTGATGAGTGTGTACGTCGTCGAGCTGCTCAGGTCTTTCTTTTATGTCACGGA 1749

Db   2841 GACCACGTTTCAAAAGAACAGGCTCTTTTTCTACCGGAAGAGTGTCTGGAGCAAGTTGCA 2900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1750 GACCACGTTTCAAAAGAACAGGCTCTTTTTCTACCGGAAGAGTGTCTGGAGCAAGTTGCA 1809

Db   2901 AAGCATTGGAATCAGACAGCACTTGAAGAGGGTGCAGCTGCGGGAGCTGTCGGAAGCAGA 2960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1810 AAGCATTGGAATCAGACAGCACTTGAAGAGGGTGCAGCTGCGGGAGCTGTCGGAAGCAGA 1869

Db   2961 GGTCAGGCAGCATCGGGAAGCCAGGCCCGCCCTGCTGACGTCCAGACTCCGCTTCATCCC 3020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1870 GGTCAGGCAGCATCGGGAAGCCAGGCCCGCCCTGCTGACGTCCAGACTCCGCTTCATCCC 1929

Db   3021 CAAGCCTGACGGGCTGCGGCCGATTGTGAACATGGACTACGTCGTGGGAGCCAGAACGTT 3080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1930 CAAGCCTGACGGGCTGCGGCCGATTGTGAACATGGACTACGTCGTGGGAGCCAGAACGTT 1989

Db   3081 CCGCAGAGAAAAGAGGGCCGAGCGTCTCACCTCGAGGGTGAAGGCACTGTTCAGCGTGCT 3140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1990 CCGCAGAGAAAAGAGGGCCGAGCGTCTCACCTCGAGGGTGAAGGCACTGTTCAGCGTGCT 2049

Db   3141 CAACTACGAGCGGGCGCGGCGCCCCGGCCTCCTGGGCGCCTCTGTGCTGGGCCTGGACGA 3200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2050 CAACTACGAGCGGGCGCGGCGCCCCGGCCTCCTGGGCGCCTCTGTGCTGGGCCTGGACGA 2109

Db   3201 TATCCACAGGGCCTGGCGCACCTTCGTGCTGCGTGTGCGGGCCCAGGACCCGCCGCCTGA 3260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2110 TATCCACAGGGCCTGGCGCACCTTCGTGCTGCGTGTGCGGGCCCAGGACCCGCCGCCTGA 2169

Db   3261 GCTGTACTTTGTCAAGGTGGATGTGACGGGCGCGTACGACACCATCCCCCAGGACAGGCT 3320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2170 GCTGTACTTTGTCAAGGTGGATGTGACGGGCGCGTACGACACCATCCCCCAGGACAGGCT 2229

Db   3321 CACGGAGGTCATCGCCAGCATCATCAAACCCCAGAACACGTACTGCGTGCGTCGGTATGC 3380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2230 CACGGAGGTCATCGCCAGCATCATCAAACCCCAGAACACGTACTGCGTGCGTCGGTATGC 2289

Db   3381 CGTGGTCCAGAAGGCCGCCCATGGGCACGTCCGCAAGGCCTTCAAGAGCCACGTCTCTAC 3440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2290 CGTGGTCCAGAAGGCCGCCCATGGGCACGTCCGCAAGGCCTTCAAGAGCCACGTCTCTAC 2349

Db   3441 CTTGACAGACCTCCAGCCGTACATGCGACAGTTCGTGGCTCACCTGCAGGAGACCAGCCC 3500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2350 CTTGACAGACCTCCAGCCGTACATGCGACAGTTCGTGGCTCACCTGCAGGAGACCAGCCC 2409
```

```
Db   3501 GCTGAGGGATGCCGTCGTCATCGAGCAGAGCTCCTCCCTGAATGAGGCCAGCAGTGGCCT 3560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2410 GCTGAGGGATGCCGTCGTCATCGAGCAGAGCTCCTCCCTGAATGAGGCCAGCAGTGGCCT 2469

Db   3561 CTTCGACGTCTTCCTACGCTTCATGTGCCACCACGCCGTGCGCATCAGGGGCAAGTCCTA 3620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2470 CTTCGACGTCTTCCTACGCTTCATGTGCCACCACGCCGTGCGCATCAGGGGCAAGTCCTA 2529

Db   3621 CGTCCAGTGCCAGGGGATCCCGCAGGGCTCCATCCTCTCCACGCTGCTCTGCAGCCTGTG 3680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2530 CGTCCAGTGCCAGGGGATCCCGCAGGGCTCCATCCTCTCCACGCTGCTCTGCAGCCTGTG 2589

Db   3681 CTACGGCGACATGGAGAACAAGCTGTTTGCGGGGATTCGGCGGGACGGGCTGCTCCTGCG 3740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2590 CTACGGCGACATGGAGAACAAGCTGTTTGCGGGGATTCGGCGGGACGGGCTGCTCCTGCG 2649

Db   3741 TTTGGTGGATGATTTCTTGTTGGTGACACCTCACCTCACCCACGCGAAAACCTTCCTCAG 3800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2650 TTTGGTGGATGATTTCTTGTTGGTGACACCTCACCTCACCCACGCGAAAACCTTCCTCAG 2709

Db   3801 GACCCTGGTCCGAGGTGTCCCTGAGTATGGCTGCGTGGTGAACTTGCGGAAGACAGTGGT 3860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2710 GACCCTGGTCCGAGGTGTCCCTGAGTATGGCTGCGTGGTGAACTTGCGGAAGACAGTGGT 2769

Db   3861 GAACTTCCCTGTAGAAGACGAGGCCCTGGGTGGCACGGCTTTTGTTCAGATGCCGGCCCA 3920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2770 GAACTTCCCTGTAGAAGACGAGGCCCTGGGTGGCACGGCTTTTGTTCAGATGCCGGCCCA 2829

Db   3921 CGGCCTATTCCCCTGGTGCGGCCTGCTGCTGGATACCCGGACCCTGGAGGTGCAGAGCGA 3980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2830 CGGCCTATTCCCCTGGTGCGGCCTGCTGCTGGATACCCGGACCCTGGAGGTGCAGAGCGA 2889

Db   3981 CTACTCCAGCTATGCCCGGACCTCCATCAGAGCCAGTCTCACCTTCAACCGCGGCTTCAA 4040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2890 CTACTCCAGCTATGCCCGGACCTCCATCAGAGCCAGTCTCACCTTCAACCGCGGCTTCAA 2949

Db   4041 GGCTGGGAGGAACATGCGTCGCAAACTCTTTGGGGTCTTGCGGCTGAAGTGTCACAGCCT 4100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2950 GGCTGGGAGGAACATGCGTCGCAAACTCTTTGGGGTCTTGCGGCTGAAGTGTCACAGCCT 3009

Db   4101 GTTTCTGGATTTGCAGGTGAACAGCCTCCAGACGGTGTGCACCAACATCTACAAGATCCT 4160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3010 GTTTCTGGATTTGCAGGTGAACAGCCTCCAGACGGTGTGCACCAACATCTACAAGATCCT 3069

Db   4161 CCTGCTGCAGGCGTACAGGTTTCACGCATGTGTGCTGCAGCTCCCATTTCATCAGCAAGT 4220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3070 CCTGCTGCAGGCGTACAGGTTTCACGCATGTGTGCTGCAGCTCCCATTTCATCAGCAAGT 3129

Db   4221 TTGGAAGAACCCCACATTTTTCCTGCGCGTCATCTCTGACACGGCCTCCCTCTGCTACTC 4280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3130 TTGGAAGAACCCCACATTTTTCCTGCGCGTCATCTCTGACACGGCCTCCCTCTGCTACTC 3189

Db   4281 CATCCTGAAAGCCAAGAACGCAGGGATGTCGCTGGGGGCCAAGGGCGCCGCCGGCCCTCT 4340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3190 CATCCTGAAAGCCAAGAACGCAGGGATGTCGCTGGGGGCCAAGGGCGCCGCCGGCCCTCT 3249

Db   4341 GCCCTCCGAGGCCGTGCAGTGGCTGTGCCACCAAGCATTCCTGCTCAAGCTGACTCGACA 4400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3250 GCCCTCCGAGGCCGTGCAGTGGCTGTGCCACCAAGCATTCCTGCTCAAGCTGACTCGACA 3309

Db   4401 CCGTGTCACCTACGTGCCACTCCTGGGGTCACTCAGGACAGCCCAGACGCAGCTGAGTCG 4460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3310 CCGTGTCACCTACGTGCCACTCCTGGGGTCACTCAGGACAGCCCAGACGCAGCTGAGTCG 3369

Db   4461 GAAGCTCCCGGGGACGACGCTGACTGCCCTGGAGGCCGCAGCCAACCCGGCACTGCCCTC 4520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3370 GAAGCTCCCGGGGACGACGCTGACTGCCCTGGAGGCCGCAGCCAACCCGGCACTGCCCTC 3429

Db   4521 AGACTTCAAGACCATCCTGGACT 4543
          |||||||||||||||||||||||
Qy   3430 AGACTTCAAGACCATCCTGGACT 3452
```

```
RESULT    4
LOCUS       AF051911     3426 bp    mRNA            ROD       02-APR-1998
DEFINITION  Mus musculus telomerase reverse transcriptase mRNA, complete cds.
ACCESSION   AF051911
NID         g3005591
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 3426)
  AUTHORS   Greenberg,R.A., Allsopp,R.C., Chin,L., Morin,G.B. and DePinho,R.A.
  TITLE     Expression of mouse telomerase reverse transcriptase during
            development, differentiation, and proliferation
  JOURNAL   Oncogene (1998) In press
REFERENCE   2  (bases 1 to 3426)
  AUTHORS   Greenberg,R.A., Allsopp,R.C., Chin,L., Morin,G.B. and DePinho,R.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-MAR-1998) Microbiology and Immunology, Albert
            Einstein College of Medicine, 1300 Morris Park Ave., Bronx, NY
            10461, USA
FEATURES             Location/Qualifiers
     source          1. .3426
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="13"
                     /map="MGD offset position 40; 40 cM from centromere"
     CDS             30. .3398
                     /note="mTERT"
                     /codon_start=1
                     /product="telomerase reverse transcriptase"
                     /db_xref="PID:g3005592"
                     /translation="MTRAPRCPAVRSLLRSRYREVWPLATFVRRLGPEGRRLVQPGDP
                     KIYRTLVAQCLVCMHWGSQPPPADLSFHQVSSLKELVARVVQRLCERNERNVLAFGFE
                     LLNEARGGPPMAFTSSVRSYLPNTVIETLRVSGAWMLLLSRVGDDLLVYLLAHCALYL
                     LVPPSCAYQVCGSPLYQICATTDIWPSVSASYRPTRPVGRNFTNLRFLQQIKSSSRQE
                     APKPLALPSRGTKRHLSLTSTSVPSAKKARCYPVPRVEEGPHRQVLPTPSGKSWVPSP
                     ARSPEVPTAEKDLSSKGKVSDLSLSGSVCCKHKPSSTSLLSPPRQNAFQLRPFIETRH
                     FLYSRGDGQERLNPSFLLSNLQPNLTGARRLVEIIFLGSRPRTSGPLCRTHRLSRRYW
                     QMRPLFQQLLVNHAECQYVRLLRSHCRFRTANQQVTDALNTSPPHLMDLLRLHSSPWQ
                     VYGFLRACLCKVVSASLWGTRHNERRFFKNLKKFISLGKYGKLSLQELMWKMKVEDCH
                     WLRSSPGKDRVPAAEHRLRERILATFLFWLMDTYVVQLLRSFFYITESTFQKNRLFFY
                     RKSVWSKLQSIGVRQHLERVRLRELSQEEVRHHQDTWLAMPICRLRFIPKPNGLRPIV
                     NMSYSMGTRALGRRKQAQHFTQRLKTLFSMLNYERTKHPHLMGSSVLGMNDIYRTWRA
                     FVLRVRALDQTPRMYFVKADVTGAYDAIPQGKLVEVVANMIRHSESTYCIRQYAVVRR
                     DSQGQVHKSFRRQVTTLSDLQPYMGQFLKHLQDSDASALRNSVVIEQSISMNESSSSL
                     FDFFLHFLRHSVVKIGDRCYTQCQGIPQGSSLSTLLCSLCFGDMENKLFAEVQRDGLL
                     LRFVDDFLLVTPHLDQAKTFLSTLVHGVPEYGCMINLQKTVVNFPVEPGTLGGAAPYQ
                     LPAHCLFPWCGLLLDTQTLEVFCDYSGYAQTSIKTSLTFQSVFKAGKTMRNKLLSVLR
                     LKCHGLFLDLQVNSLQTVCINIYKIFLLQAYRFHACVIQLPFDQRVRKNLTFFLGIIS
                     SQASCCYAILKVKNPGMTLKASGSFPPEAAHWLCYQAFLLKLAAHSVIYKCLLGPLRT
                     AQKLLCRKLPEATMTILKAAADPALSTDFQTILD"
BASE COUNT      746 a    979 c    907 g    794 t
ORIGIN
```

```
  Query Match           22.9%;  Score 920;  DB 22;  Length 3426;
  Best Local Similarity  69.2%;  Pred. No. 0.00e+00;
  Matches  2379;  Conservative    0;  Mismatches 979;  Indels  80;  Gaps  42;

Db     14 CCCGGCCTTGAGCACAATGACCCGCGCTCCTCGTTGCCCCGCGGTGCGCTCTCTGCTGCG 73
          |||||||     | | ||| | ||||||||  || ||||  || ||||||||  ||||||||
Qy     40 CCCGGCCACCCCCGCGATGCCGCGCGCTCCCCGCTGCCGAGCCGTGCGCTCCCTGCTGCG 99

Db     74 CAGCCGATACCGGGAGGTGTGGCCGCTGGCAACCTTTGTGCGGCGCCTGGGGCCCGAGGG 133
          |||||  |||||  ||||||  |||||||||  ||  ||  ||||||||||||||||  ||||
Qy    100 CAGCCACTACCGCGAGGTGCTGCCGCTGGCCACGTTCGTGCGGCGCCTGGGGCCCCAGGG 159

Db    134 CAGGCGGCTTGTGCAACCCGGGGACCCGAAGATCTACCGCACTTTGGTTGCCCAATGCCT 193
          | |||||||  |||||  | ||||||||||  |   | ||||  |  ||||  |||||  |||||
Qy    160 CTGGCGGCTGGTGCAGCGCGGGGACCCGGCGGCTTTCCGCGCGCTGGTGGCCCAGTGCCT 219

Db    194 AGTGTGCATGCACTGGGGCTCACAGCCTCCACCTGCCGACCTTTCCTTCCACCAGGTGTC 253
          ||||||  |||  |||||  |  |||  |||  ||  ||  ||||  ||  |||||||  |||||||||
```

```
Qy    220 GGTGTGCGTGCCCTGGGACGCACGGCCGCCCCCCGCCGCCCCCTCCTTCCGCCAGGTGTC 279
Db    254 ATCCCTGAAAGAGCTGGTGGCCAGGGTTGTGCAGAGACTCTGCGAGCGCAACGAGAGAAA 313
Qy    280 CTGCCTGAAGGAGCTGGTGGCCCGAGTGCTGCAGAGGCTGTGCGAGCGCGGCGCGAAGAA 339
Db    314 CGTGCTGGCTTTTGGCTTTGAGCTGCTTAACGAGGCCAGAGGCGGGCCTCCCATGGCCTT 373
Qy    340 CGTGCTGGCCTTCGGCTTCGCGCTGCTGGACGGGGCCCGCGGGGGCCCCCCCGAGGCCTT 399
Db    374 CACTAGTAGCGTGCGTAGCTACTTGCCCAACACTGTTATTGAGACCCTGCGTGTCAGTGG 433
Qy    400 CACCACCAGCGTGCGCAGCTACCTGCCCAACACGGTGACCGACGCACTGCGGGGGAGCGG 459
Db    434 TGCATGGATGCTACTGTTGAGCCGAGTGGGCGACGACCTGCTGGTCTACCTGCTGGCACA 493
Qy    460 GGCGTGGGGGCTGCTGCTGCGCCGCGTGGGCGACGACGTGCTGGTTCACCTGCTGGCACG 519
Db    494 CTGTGCTCTTTATCTTCTGGTGCCCCCCAGCTGTGCCTACCAGGTGTGTGGGTCTCCCCT 553
Qy    520 CTGCGCGCTCTTTGTGCTGGTGGCTCCCAGCTGCGCCTACCAGGTGTGCGGGCCGCCGCT 579
Db    554 GTACCAAATTTGTGCCACCACGGATATCTGGCCCTCTGTGTCCGCTAGTTACAGGCCCAC 613
Qy    580 GTACCAGCTCGGCGCTGCCACTCAGGCCCGGCCCCCGCCACACGCTAGTGGACCCCGAAG 639
Db    614 CCGACCCGTGGGCAGGAATTTCACTAACCTTAGGTTCTTACAACAGATCAAGAGCAGTAG 673
Qy    640 GCGTCTGGGATGCGAACGGGCCTGGAACCATAGCGTC--AGGG-AGGCCGGGGTCCCCCT 696
Db    674 TCGCCAGGAAGCACCGAAACCCCTGGCCTTGCCATCTCGAGGTACAAAGAGGCATCTGAG 733
Qy    697 GGGCCTGCCAGCCCCGGGTGCGAGGAGGCGCGGGGGCAGTGCCAGCCGAAGTCTGCCGTT 756
Db    734 TCTCACCAGTACAAGTGTGCCTTCAGCTAAGAAGGCCAGATGCTATCCTGTCCCGAGAGT 793
Qy    757 GCCCAAGAGGCCCAGGCGTGGCGCTGCCCCTGAGCCGGAGCGGACGCCCGTTGGGC-AGG 815
Db    794 GGAGGAGGGACCCCACAGGCAGGTGCTACCAACCCCATCAG-GCAAATCATGGGTGCCAA 852
Qy    816 GGTCCTGGGCCCACCCGGGCAGGACGCGTGGACCGAGTGACCGTGGTTTCTGTGTGGTGT 875
Db    853 GTCCTGCTCGGTCCCCCGAGGTGCCTACTGCAGAGAAAGATTTGTCTTCTAAAGGAAAGG 912
Qy    876 CACCTGCCAGACCCGCCGAAGAAGCCACCTCTTTGGAGGGTGCG-CT-CTCT-GGCACGC 932
Db    913 TGTCTGACCTGAGTCTCTCTGGGTCGGTGTGCTGTAAACACAAGCCCAGCTCCACATCTC 972
Qy    933 -GCCACTCCCACCCATCCGTGGGCCGCCA-GCACCACGCGGGCCCCCCA-TCCACATCGC 989
Db    973 TGCTGTCACCACCCCGCCAAAATGCCTTTCAGCTCAGGCCATTTATTGAGACCAGACATT 1032
Qy    990 GGCCACCACGTCCCTGGGACAC-GCCTTGTCCCCC-GGTG-TACGCCGAGACCAAGCACT 1046
Db   1033 TCCTTTACTCCAGGGGAGATGGCCAAGAGCGTCTAAACCCCTCATTCCTACTCAGCAACC 1092
Qy   1047 TCCTCTACTCCTCAGGCGA---CAAGGAGCAGCTGCGGCCCTCCTTCCTACTCAGCTCTC 1103
Db   1093 TCCAGCCTAACTTGACTGGGGCCAGGAGACTGGTGGAGATCATCTTTCTGGGCTCAAGGC 1152
Qy   1104 TGAGGCCCAGCCTGACTGGCGCTCGGAGGCTCGTGGAGACCATCTTTCTGGGTTCCAGGC 1163
Db   1153 CTAGGACATCAGGACCACTCTGCAGGACACACCGTCTATCGCGTCGATACTGGCAGATGC 1212
Qy   1164 CCTGGATGCCAGGGACTCCCCGCAGGTTGCCCCGCCTGCCCCAGCGCTACTGGCAAATGC 1223
Db   1213 GGCCCCTGTTCCAACAGCTGCTGGTGAACCATGCAGAGTGCCAATA-----TG----TCA 1263
Qy   1224 GGCCCCTGTTTCTGGAGCTGCTTGGGAACCACGCGCAGTGCCCCTACGGGGTGCTCCTCA 1283
Db   1264 ----G-ACT---C-CT-C-AG-----GTCACATT-GCAG---GTTTC---GAACAGCA-A 1299
Qy   1284 AGACGCACTGCCCGCTGCGAGCTGCGGTCACCCCAGCAGCCGGTGTCTGTGCCCGGGAGA 1343

Db   1300 ACCAACAGG------TGACAGATGCCTTG-A--AC-AC-CAGCCCACCG-CACCTCATGG 1347
Qy   1344 AGCCCCAGGGCTCTGTGGCGGCCCCCGAGGAGGAGGACACAGACCCCCGTCGCCTGGTGC 1403
Db   1348 ATTTGCTCCGCCTGCACAGCAGTCCCTGGCAGGTATATGGTTTTCTTCGGGCCTGTCTCT 1407
Qy   1404 AGCTGCTCCGCCAGCACAGCAGCCCCTGGCAGGTGTACGGCTTCGTGCGGGCCTGCCTGC 1463
Db   1408 GCAAGGTGGTGTCTGCTAGTCTCTGGGGTACCAGGCACAATGAGCGCCGCTTCTTTAAGA 1467
Qy   1464 GCCGGCTGGTGCCCCCAGGCCTCTGGGGCTCCAGGCACAACGAACGCCGCTTCCTCAGGA 1523
Db   1468 ACTTAAAGAAGTTCATCTCGTTGGGGAAATACGGCAAGCTATCACTGCAGGAACTGATGT 1527
Qy   1524 ACACCAAGAAGTTCATCTCCCTGGGGAAGCATGCCAAGCTCTCGCTGCAGGAGCTGACGT 1583
Db   1528 GGAAGATGAAAGTAGAGGATTGCCACTGGCTCCGCAGCAGCCCGGGGAAGGACCGTGTCC 1587
Qy   1584 GGAAGATGAGCGTGCGGGACTGCGCTTGGCTGCGCAGGAGCCCAGGGGTTGGCTGTGTTC 1643
Db   1588 CCGCTGCAGAGCACCGTCTGAGGGAGAGGATCCTGGCTACGTTCCTGTTCTGGCTGATGG 1647
Qy   1644 CGGCCGCAGAGCACCGTCTGCGTGAGGAGATCCTGGCCAAGTTCCTGCACTGGCTGATGA 1703
Db   1648 ACACATACGTGGTACAGCTGCTTAGGTCATTCTTTTACATCACAGAGAGCACATTCCAGA 1707
Qy   1704 GTGTGTACGTCGTCGAGCTGCTCAGGTCTTTCTTTTATGTCACGGAGACCACGTTTCAAA 1763
Db   1708 AGAACAGGCTCTTCTTCTACCGTAAGAGTGTGTGGAGCAAGCTGCAGAGCATTGGAGTCA 1767
Qy   1764 AGAACAGGCTCTTTTTCTACCGGAAGAGTGTCTGGAGCAAGTTGCAAAGCATTGGAATCA 1823
Db   1768 GGCAACACCTTGAGAGAGTGCGGCTACGGGAGCTGTCACAAGAGGAGGTCAGGCATCACC 1827
Qy   1824 GACAGCACTTGAAGAGGGTGCAGCTGCGGGAGCTGTCGGAAGCAGAGGTCAGGCAGCATC 1883
Db   1828 AGGACACCTGGCTAGCCATGCCCATCTGCAGACTGCGCTTCATCCCCAAGCCCAACGGCC 1887
Qy   1884 GGGAAGCCAGGCCCGCCCTGCTGACGTCCAGACTCCGCTTCATCCCCAAGCCTGACGGGC 1943
Db   1888 TGCGGCCCATTGTGAACATGAGTTATAGCATGGGTACCAGAGCTTTGGGCAGAAGGAAGC 1947
Qy   1944 TGCGGCCGATTGTGAACATGGACTACGTCGTGGGAGCCAGAACGTTCCGCAGAGAAAAGA 2003
Db   1948 AGGCCCAGCATTTCACCCAGCGTCTCAAGACTCTCTTCAGCATGCTCAACTATGAGCGGA 2007
Qy   2004 GGGCCGAGCGTCTCACCTCGAGGGTGAAGGCACTGTTCAGCGTGCTCAACTACGAGCGGG 2063
Db   2008 CAAAACATCCTCACCTTATGGGGTCTTCTGTACTGGGTATGAATGACATCTACAGGACCT 2067
Qy   2064 CGCGGCGCCCCGGCCTCCTGGGCGCCTCTGTGCTGGGCCTGGACGATATCCACAGGGCCT 2123
Db   2068 GGCGGGCCTTTGTGCTGCGTGTGCGTGCTCTGGACCAGACACCCAGGATGTACTTTGTTA 2127
Qy   2124 GGCGCACCTTCGTGCTGCGTGTGCGGGCCCAGGACCCGCCGCCTGAGCTGTACTTTGTCA 2183
Db   2128 AGGCAGATGTGACCGGGGCCTATGATGCCATCCCCCAGGGTAAGCTGGTGGAGGTTGTTG 2187
Qy   2184 AGGTGGATGTGACGGGCGCGTACGACACCATCCCCCAGGACAGGCTCACGGAGGTCATCG 2243
Db   2188 CCAATATGATCAGGCACTCGGAGAGCACGTACTGTATCCGCCAGTATGCAGTGGTCCGGA 2247
Qy   2244 CCAGCATCATCAAACCC-CAGA-A-CACGTACTGCGTGCGTCGGTATGCCGTGGTCCAGA 2300
Db   2248 GAGATAGCCAAGGCCAAGTCCACAAGTCCTTTAGGAGACAGGTCACCACCCTCTCTGACC 2307
Qy   2301 AGGCCGCCCATGGGCACGTCCGCAAGGCCTTCAAGAGCCACGTCTCTACCTTGACAGACC 2360
Db   2308 TCCAGCCATACATGGGCCAGTTCCTTAAGCATCTGCAGGATTCAGATGCCAGTGCACTGA 2367
Qy   2361 TCCAGCCGTACATGCGACAGTTCGTGGCTCACCTGCAGGA----GA--CCAGCCCGCTGA 2414
```

```
Db    2368 GGAACTCCGTTGTCATCGAGCAGAGCATCTCTATGAATGAGAGCAGCAGCAGCCTGTTTG 2427
Qy    2415 GGGATGCCGTCGTCATCGAGCAGAGCTCCTCCCTGAATGAGGCCAGCAGTGGCCTCTTCG 2474

Db    2428 ACTTCTTCCTGCACTTCCTGCGTCACAGTGTCGTAAAGATTGGTGACAGGTGCTATACGC 2487
Qy    2475 ACGTCTTCCTACGCTTCATGTGCCACCACGCCGTGCGCATCAGGGGCAAGTCCTACGTCC 2534

Db    2488 AGTGCCAGGGCATCCCCCAGGGCTCCAGCCTATCCACCCTGCTCTGCAGTCTGTGTTTCG 2547
Qy    2535 AGTGCCAGGGGATCCCGCAGGGCTCCATCCTCTCCACGCTGCTCTGCAGCCTGTGCTACG 2594

Db    2548 GAGACATGGAGAACAAGCTGTTTGCTGAGGTGCAGCGGGATGGGTTGCTTTTACGTTTTG 2607
Qy    2595 GCGACATGGAGAACAAGCTGTTTGCGGGGATTCGGCGGGACGGGCTGCTCCTGCGTTTGG 2654

Db    2608 TTGATGACTTTCTGTTGGTGACGCCTCACTTGGACCAAGCAAAAACCTTCCTCAGCACCC 2667
Qy    2655 TGGATGATTTCTTGTTGGTGACACCTCACCTCACCCACGCGAAAACCTTCCTCAGGACCC 2714

Db    2668 TGGTCCATGGCGTTCCTGAGTATGGGTGCATGATAAACTTGCAGAAGACAGTGGTGAACT 2727
Qy    2715 TGGTCCGAGGTGTCCCTGAGTATGGCTGCGTGGTGAACTTGCGGAAGACAGTGGTGAACT 2774

Db    2728 TCCCTGTGGAGCCTGGTACCCTGGGTGGTGCAGCTCCATACCAGCTGCCTGCTCACTGCC 2787
Qy    2775 TCCCTGTAGAAGACGAGGCCCTGGGTGGCACGGCTTTTGTTCAGATGCCGGCCCACGGCC 2834

Db    2788 TGTTTCCCTGGTGTGGCTTGCTGCTGGACACTCAGACTTTGGAGGTGTTCTGTGACTACT 2847
Qy    2835 TATTCCCCTGGTGCGGCCTGCTGCTGGATACCCGGACCCTGGAGGTGCAGAGCGACTACT 2894

Db    2848 CAGGTTATGCCCAGACCTCAATTAAGACGAGCCTCACCTTCCAGAGTGTCTTCAAAGCTG 2907
Qy    2895 CCAGCTATGCCCGGACCTCCATCAGAGCCAGTCTCACCTTCAACCGCGGCTTCAAGGCTG 2954

Db    2908 GGAAGACCATGCGGAACAAGCTCCTGTCGGTCTTGCGGTTGAAGTGTCACGGTCTATTTC 2967
Qy    2955 GGAGGAACATGCGTCGCAAACTCTTTGGGGTCTTGCGGCTGAAGTGTCACAGCCTGTTTC 3014

Db    2968 TAGACTTGCAGGTGAACAGCCTCCAGACAGTCTGCATCAATATATACAAGATCTTCCTGC 3027
Qy    3015 TGGATTTGCAGGTGAACAGCCTCCAGACGGTGTGCACCAACATCTACAAGATCCTCCTGC 3074

Db    3028 TTCAGGCCTACAGGTTCCATGCATGTGTGATTCAGCTTCCCTTTGACCAGCGTGTTAGGA 3087
Qy    3075 TGCAGGCGTACAGGTTTCACGCATGTGTGCTGCAGCTCCCATTTCATCAGCAAGTTTGGA 3134

Db    3088 AGAACCTCACATTCTTTCTGGGCATCATCTCCAGCCAAGCATCCTGCTGCTATGCTATCC 3147
Qy    3135 AGAACCCCACATTTTTCCTGCGCGTCATCTCTGACACGGCCTCCCTCTGCTACTCCATCC 3194

Db    3148 TGAAGGTCAAGAATCCAGGAATGACACTAAAGGCCTCTGGCTCCTTT----C-CT--CCT 3200
Qy    3195 TGAAAGCCAAGAACGCAGGGATGTCGCTGGGGGCCAAGGGCGCCGCCGGCCCTCTGCCCT 3254

Db    3201 --GAAGCCGCACATTGGCTCTGCTACCAGGCCTTCCTGCTCAAGCTGGCTGCTCATTCTG 3258
Qy    3255 CCGAGGCCGTGCAGTGGCTGTGCCACCAAGCATTCCTGCTCAAGCTGACTCGACACCGTG 3314

Db    3259 TCATCTACAAATGTCTCCTGGGACCTCTGAGGACAGCCCAAAAACTGCTGTGCCGGAAGC 3318
Qy    3315 TCACCTACGTGCCACTCCTGGGGTCACTCAGGACAGCCCAGACGCAGCTGAGTCGGAAGC 3374

Db    3319 TCCCAGAGGCGACAATGACCATCCTTAAAGCTGCAGCTGACCCAGCCCTAAGCACAGACT 3378
Qy    3375 TCCCGGGGACGACGCTGACTGCCCTGGAGGCCGCAGCCAACCCGGCACTGCCCTCAGACT 3434

Db    3379 TTCAGACCATTTTGGACT 3396
Qy    3435 TCAAGACCATCCTGGACT 3452
```

```
RESULT    5
LOCUS       AF029235      201 bp    mRNA            ROD       12-NOV-1997
DEFINITION  Mus musculus telomerase catalytic subunit mRNA, partial cds.
ACCESSION   AF029235
NID         g2605902
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 201)
  AUTHORS   Drissi,R. and Cleveland,J.L.
  TITLE     Partial sequence of Mus musculus telomerase catalytic subunit
            homolog
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 201)
  AUTHORS   Drissi,R. and Cleveland,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1997) Biochemistry, St jude Children's Research
            Hospital, 332 North Lauderdale, Memphis, TN 38105, USA
FEATURES             Location/Qualifiers
     source          1. .201
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /cell_line="32D.3"
     CDS             <1. .>201
                     /codon_start=1
                     /product="telomerase catalytic subunit"
                     /db_xref="PID:g2605903"
                     /translation="FFYVTESTFQKNRLFFYRKSVWSKLQSIGVRQHLERVRLRELSQ
                     EEVRHHQDTWLAMPICRLRFIPK"
BASE COUNT       51 a     54 c     55 g     41 t
ORIGIN
```

```
  Query Match             3.2%;  Score 129;  DB 22;  Length 201;
  Best Local Similarity  82.1%;  Pred. No. 2.00e-73;
  Matches   165;  Conservative    0;  Mismatches  36;  Indels   0;  Gaps   0;

Db       1 TTCTTTTATGTCACGGAGAGCACATTCCAGAAGAACAGGCTCTTCTTCTACCGTAAGAGT 60
Qy    1733 TTCTTTTATGTCACGGAGACCACGTTTCAAAAGAACAGGCTCTTTTTCTACCGGAAGAGT 1792

Db      61 GTGTGGAGCAAGCTGCAGAGCATTGGAGTCAGGCAACACCTTGAGAGAGTGCGGCTACGG 120
Qy    1793 GTCTGGAGCAAGTTGCAAAGCATTGGAATCAGACAGCACTTGAAGAGGGTGCAGCTGCGG 1852

Db     121 GAGCTGTCACAAGAGGAGGTCAGGCATCACCAGGACACCTGGCTAGCCATGCCCATCTGC 180
Qy    1853 GAGCTGTCGGAAGCAGAGGTCAGGCAGCATCGGGAAGCCAGGCCCGCCCTGCTGACGTCC 1912

Db     181 AGACTCCGCTTCATCCCCAAG 201
Qy    1913 AGACTCCGCTTCATCCCCAAG 1933
```

```
RESULT    6
LOCUS       I66494       7218 bp    DNA             PAT       23-DEC-1997
DEFINITION  Sequence 14 from patent US 5670367.
ACCESSION   I66494
NID         g2724471
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7218)
  AUTHORS   Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
  TITLE     Recombinant fowlpox virus
  JOURNAL   Patent: US 5670367-A 14 23-SEP-1997;
FEATURES             Location/Qualifiers
     source          1. .7218
                     /organism="unknown"
BASE COUNT     1944 a   1491 c   1486 g   1929 t    368 others
ORIGIN
```

```
  Query Match            1.7%;  Score 67;  DB 17;  Length 7218;
  Best Local Similarity  2.4%;  Pred. No. 3.56e-26;
  Matches     9;  Conservative  211;  Mismatches 153;  Indels   0;  Gaps   0;

Db   1055 GGAGCTTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1114
          ||||  |||| |:::: : : : :: ::::: ::: ::: :   :: ::  : :: : :
Qy    910 GGAGGGTGCGCTCTCTGGCACGCGCCACTCCCACCCATCCGTGGGCCGCCAGCACCACGC 969

Db   1115 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1174
             :::::: ::: : :: :  :: :: : :::::    : : :::: ::::::  : :
Qy    970 GGGCCCCCCATCCACATCGCGGCCACCACGTCCCTGGGACACGCCTTGTCCCCCGGTGTA 1029

Db   1175 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1234
          : ::    ::   : :::::::: :::::::   :  :      :  :: : :::::::::
Qy   1030 CGCCGAGACCAAGCACTTCCTCTACTCCTCAGGCGACAAGGAGCAGCTGCGGCCCTCCTT 1089

Db   1235 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1294
          ::: :::  :::::::    :::  :::  ::  : :::      ::: :      :: ::::
Qy   1090 CCTACTCAGCTCTCTGAGGCCCAGCCTGACTGGCGCTCGGAGGCTCGTGGAGACCATCTT 1149

Db   1295 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1354
          :::   ::::   ::::   : ::      :::::: :   :: :::: ::: ::::  :
Qy   1150 TCTGGGTTCCAGGCCCTGGATGCCAGGGACTCCCCGCAGGTTGCCCCGCCTGCCCCAGCG 1209

Db   1355 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1414
          :: ::  :   : :  ::::: :::::    :: :::     :: : : :  : :::::
Qy   1210 CTACTGGCAAATGCGGCCCCTGTTTCTGGAGCTGCTTGGGAACCACGCGCAGTGCCCCTA 1269

Db   1415 YYYYYYYYYYYYY 1427
          :    : ::::::
Qy   1270 CGGGGTGCTCCTC 1282


RESULT    7
LOCUS       I66494       7218 bp    DNA             PAT       23-DEC-1997
DEFINITION  Sequence 14 from patent US 5670367.
ACCESSION   I66494
NID         g2724471
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7218)
  AUTHORS   Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
  TITLE     Recombinant fowlpox virus
  JOURNAL   Patent: US 5670367-A 14 23-SEP-1997;
FEATURES              Location/Qualifiers
     source           1. .7218
                      /organism="unknown"
BASE COUNT     1944 a   1491 c   1486 g   1929 t    368 others
ORIGIN

  Query Match            1.2%;  Score 49;  DB 17;  Length 7218;
  Best Local Similarity  2.0%;  Pred. No. 1.15e-13;
  Matches     6;  Conservative  165;  Mismatches 122;  Indels   0;  Gaps   0;

Db   1150 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1209
          :: : ::: ::: :  ::   : : ::   :  : ::::: :::::    :    : :::
Cp   2068 CCGCGCCCGCTCGTAGTTGAGCACGCTGAACAGTGCCTTCACCCTCGAGGTGAGACGCTC 2009

Db   1210 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1269
            :::::::::::::: :    : ::::  ::::: :  : :  ::: : ::: :  ::
Cp   2008 GGCCCTCTTTTCTCTGCGGAACGTTCTGGCTCCCACGACGTAGTCCATGTTCACAATCGG 1949

Db   1270 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1329
          :: :  ::: ::   :::      :    :    :::   : ::  :    :   ::: :
Cp   1948 CCGCAGCCCGTCAGGCTTGGGGATGAAGCGGAGTCTGGACGTCAGCAGGGCGGGCCTGGC 1889

Db   1330 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1389
          :::::  : :: :::  ::::: ::::::  :  ::::: :  :: : :::::::    :
Cp   1888 TTCCCGATGCTGCCTGACCTCTGCTTCCGACAGCTCCCGCAGCTGCACCCTCTTCAAGTG 1829

Db   1390 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYGTACCAAA 1442
          :: :::  ::::  : :::: :  ::: ::::    : :::::::: |||  |||
Cp   1828 CTGTCTGATTCCAATGCTTTGCAACTTGCTCCAGACACTCTTCCGGTAGAAAA 1776


RESULT    8
LOCUS       I28278       215 bp    DNA             PAT       30-OCT-1996
DEFINITION  Sequence 5 from patent US 5569830.
ACCESSION   I28278
NID         g1819054
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 215)
  AUTHORS   Bennett,A., Labavitch,J.M., Powell,A. and Stotz,H.
  TITLE     Plant inhibitors of fungal polygalacturonases and their use to
            control fungal disease
  JOURNAL   Patent: US 5569830-A 5 29-OCT-1996;
FEATURES              Location/Qualifiers
     source           1. .215
                      /organism="unknown"
BASE COUNT       15 a      8 c     25 g     26 t    141 others
ORIGIN

  Query Match            1.0%;  Score 42;  DB 17;  Length 215;
  Best Local Similarity 15.5%;  Pred. No. 4.08e-09;
  Matches    32;  Conservative   82;  Mismatches  90;  Indels   2;  Gaps   2;

Db      7 SSSSVVSRTASCNDKAKKDGNTTSSWTTDCCNRTWGVCDTDTTYRVNNDSGHNKYSSANY 66
          :::::::: :  :  ::  || : | :   :|: : : :|   :  :: :  ::   :
Qy   2629 GCGGGACGGGCTGCTCCTGCGTTTGGTGGATGATTTCTTGTTGGTGACACCTCACCTCAC 2688

Db     67 NYGGNNVGAAKTHYYTHTNVSGADSKTVTDSYNASGTSSSNGGTDGNR-SGADSYGSSKT 125
          :    : ||  ::: :| ::|| : |: ::: |:||: :    :| : :| :::| ::|
Qy   2689 CCACGCGAAAACCTTCCTCAGGACCCTGGTCCGAGGTGTCCCTGAGTATGGCTGCGTGGT 2748

Db    126 AMTSRNRTGKTANNAVDSRNMGDASVGSD-KNTKKHAKNSADGKVGSKNNGDRNNRYGTG 184
          : :  : |: |   :: :  |:|:  :  : | ::|: : :|:: ::  |::  ::|
Qy   2749 GAACTTGCGGAAGACAGTGGTGAACTTCCCTGTAGAAGACGAGGCCCTGGGTGGCACGGC 2808

Db    185 TKSNVSNNCGGGNKRDVSSYANNKCC 210
          |:  :    |     :::::|  :||
Qy   2809 TTTTGTTCAGATGCCGGCCCACGGCC 2834


RESULT    9
LOCUS       I28278       215 bp    DNA             PAT       30-OCT-1996
DEFINITION  Sequence 5 from patent US 5569830.
ACCESSION   I28278
NID         g1819054
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 215)
  AUTHORS   Bennett,A., Labavitch,J.M., Powell,A. and Stotz,H.
  TITLE     Plant inhibitors of fungal polygalacturonases and their use to
            control fungal disease
  JOURNAL   Patent: US 5569830-A 5 29-OCT-1996;
FEATURES              Location/Qualifiers
     source           1. .215
                      /organism="unknown"
BASE COUNT       15 a      8 c     25 g     26 t    141 others
ORIGIN

  Query Match            0.9%;  Score 37;  DB 17;  Length 215;
  Best Local Similarity 14.2%;  Pred. No. 5.00e-06;
  Matches    22;  Conservative   64;  Mismatches  68;  Indels   1;  Gaps   1;

Db     50 YRVNNDSGHNKYSSANYNYGGNNVGAAKTHYYTHTNVSGADSKTVTDSYNASGTSSSNGG 109
          :::  :: : :: :  :  |   :  |:  :: :  ::||::: : :::   |  ::
Cp   1128 CGAGCGCCAGTCAGGCTGGGCCTCAGAGAGCTGAGTAGGAAGGAGGGCCGCAGCTGCTCC 1069
```

```
Db    110 TDGNRSGADSYGSSKTA-MTSRNRTGKTANNAVDSRNMGDASVGSDKNTKKHAKNSADGK 168
          |:|  :  :: |: :|| : :: :|| |    :::: :|:|:: :::  : :|  : : :
Cp   1068 TTGTCGCCTGAGGAGTAGAGGAAGTGCTTGGTCTCGGCGTACACCGGGGGACAAGGCGTG 1009

Db    169 VGSKNNGDRNNRYGTGTKSNVSNNCGGGNKRDVSS 203
            :   |::    |||  : ::    || ::::::
Cp   1008 TCCCAGGGACGTGGTGGCCGCGATGTGGATGGGGG 974


RESULT   10
LOCUS       E04076        565 bp    RNA             PAT       26-NOV-1996
DEFINITION  gDNA encoding envelope region of type C hepatitis virus.
ACCESSION   E04076
NID         g2172286
KEYWORDS    JP 1992349885-A/1.
SOURCE      Hepatitis C virus.
  ORGANISM  Hepatitis C virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepatitis C-like viruses.
REFERENCE   1  (bases 1 to 565)
  AUTHORS   Tsutae,M., Kazuaki,C., Hiromitsu,K. and Yataro,I. .
  TITLE     NUCLEIC ACID FRAGMENT WITHIN ENVELOPE REGION OF HEPATITIS C VIRUS
            AND METHOD FOR DETECTING THE SAME
  JOURNAL   Patent: JP 1992349885-A 1 04-DEC-1992;
            TEIJIN LTD
COMMENT     OS   Hepatitis C virus
            PN   JP 1992349885-A/1
            PD   04-DEC-1992
            PF   29-MAY-1991 JP 1991152169
            PI   MORINAGA TSUTAE, CHAYAMA KAZUAKI, KUMADA HIROMITSU, PI
            ICHIKAWA YATARO
            PC   C12N15/10,C12Q1/68,C12Q1/70//C12N15/11;
            CC   strandedness: Single;
            CC   topology: Linear;.
FEATURES              Location/Qualifiers
     source           1. .565
                      /organism="Hepatitis C virus"
                      /db_xref="taxon:11103"
BASE COUNT       60 a     93 c    107 g     85 t    220 others
ORIGIN

  Query Match             0.9%;  Score 36;  DB 17;  Length 565;
  Best Local Similarity  31.7%;  Pred. No. 1.98e-05;
  Matches    59;  Conservative   58;  Mismatches  66;  Indels   3;  Gaps   3;

Db    233 GYRYBCCYTGYGTYCGSGARRRYRRYNNYTCBMGNTGYTGGGYRGCGCTBACYCCCACGS 292
          |: ::| :||:    |:  : ::::    |     ||: || : |||||:  :  | | :
Qy   1561 GCTCTCGCTGCAGGAGCTGACGTGGAAGATGAGCGTGCGGGACTGCGCTTGGCTGCGCAG 1620

Db    293 TYGCVRCYAGRRAYVBYAVYVTYCCCRCBRYGVMRMTWCGBCGBCAYRTCGAYYTGCTYG 352
            ||: |  |:: :::: :::|:|| :|::: :::: ||:| :| ::  ||  | || |
Qy   1621 GAGCC-CAGGGGTTGGCTGTGTTCCGGCCGCAGAGCACCGTCTGCGTGAGGAGATCCTGG 1679

Db    353 THGGGR-SVGCYRCYYTCTGYTCSGCYVTSTAYGTKGGRGAYYTBTGCGGRTCYGTYTTY 411
            :  |  ::  ::|:  ||| |  | ::|:||:|| |  || :|:  | |:||: |:||:
Qy   1680 CCAAGTTCCTGCACTGGCTGATGAG-TGTGTACGTCGTCGAGCTGCTCAGGTCTTTCTTT 1738

Db    412 CTYRTY 417
            ::|:
Qy   1739 TATGTC 1744


RESULT   11
LOCUS       A10161        201 bp    DNA             PAT       01-DEC-1993
DEFINITION  Synthetic DNA for preproprotease leader & prochymosin.
ACCESSION   A10161
NID         g490655
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 201)
  AUTHORS   Garvin,R.T. and James,E.
  TITLE     Production of active proteins containing cystine residues
  JOURNAL   Patent: EP 0222279-A 5 20-MAY-1987;
            Cangene Corporation
FEATURES              Location/Qualifiers
     source           1. .201
                      /organism="Artificial sequences"
     CDS              1. .>201
                      /codon_start=1
                      /product="preproprotease leader & prochymosin"
                      /db_xref="PID:g490656"
                      /translation="MPHSPVSPAESPAPQPGRPRPVVSRRLLEGGAAVLGALALSASP
                      LTAQAAVRRAAREITRIPLYKGK"
BASE COUNT       16 a     66 c     51 g     22 t     46 others
ORIGIN

  Query Match             0.7%;  Score 30;  DB 17;  Length 201;
  Best Local Similarity  36.1%;  Pred. No. 5.17e-02;
  Matches    52;  Conservative   35;  Mismatches  57;  Indels   0;  Gaps   0;

Db      2 TGCCSCACTCSCCSGTSTCSCCSGCSGAGTCSCCSGCSCCSCAGCCSGGSCGSCCSCGSC 61
          ||| :| | |:  :|   |:| :||:  | |: |: |: |:  |||: |:   ||:||:
Qy     15 TGCTGCGCACGTGGGAAGCCCTGGCCCCGGCCACCCCCGCGATGCCGCGCGCTCCCCGCT 74

Db     62 CSGTSGTSTCSCGSCGSCTSCTSGAGGGSGGSGCSGCSGTSCTSGGGGCCCTSGCSCTST 121
            :  | :  :||:  :||:||:    |:  :  :  :| : |:  | | ||:||:  :|
Qy     75 GCCGAGCCGTGCGCTCCCTGCTGCGCAGCCACTACCGCGAGGTGCTGCCGCTGGCCACGT 134

Db    122 CSGCSTCSCCSCTSACGGCCCAGG 145
            :| :  :| :||:  | ||||||
Qy    135 TCGTGCGGCGCCTGGGGCCCCAGG 158


RESULT   12
LOCUS       A10158        201 bp    DNA             PAT       01-DEC-1993
DEFINITION  S.griseus gene for preproprotease (partial).
ACCESSION   A10158
NID         g490196
KEYWORDS    .
SOURCE      Streptomyces griseus.
  ORGANISM  Streptomyces griseus
            Eubacteria; Firmicutes; Actinomycetes; Streptomycetes;
            Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 201)
  AUTHORS   Garvin,R.T. and James,E.
  TITLE     Production of active proteins containing cystine residues
  JOURNAL   Patent: EP 0222279-A 2 20-MAY-1987;
            Cangene Corporation
FEATURES              Location/Qualifiers
     source           1. .201
                      /organism="Streptomyces griseus"
                      /db_xref="taxon:1911"
     CDS              1. .>201
                      /codon_start=1
                      /transl_table=11
                      /product="preproprotease"
                      /db_xref="PID:g490197"
                      /translation="MPHSPVSPAESPAPQPGRPRPVVSRRLLEGGAAVLGALALSASP
                      LTAQAAVRRAAADEPPEWNDFGV"
BASE COUNT       13 a     66 c     54 g     22 t     46 others
ORIGIN

  Query Match             0.7%;  Score 30;  DB 17;  Length 201;
  Best Local Similarity  36.1%;  Pred. No. 5.17e-02;
  Matches    52;  Conservative   35;  Mismatches  57;  Indels   0;  Gaps   0;

Db      2 TGCCSCACTCSCCSGTSTCSCCSGCSGAGTCSCCSGCSCCSCAGCCSGGSCGSCCSCGSC 61
          ||| :| | |:  :|   |:| :||:  | |: |: |: |:  |||: |:   ||:||:
Qy     15 TGCTGCGCACGTGGGAAGCCCTGGCCCCGGCCACCCCCGCGATGCCGCGCGCTCCCCGCT 74

Db     62 CSGTSGTSTCSCGSCGSCTSCTSGAGGGSGGSGCSGCSGTSCTSGGGGCCCTSGCSCTST 121
            :  | :  :||:  :||:||:    |:  :  :  :| : |:  | | ||:||:  :|
Qy     75 GCCGAGCCGTGCGCTCCCTGCTGCGCAGCCACTACCGCGAGGTGCTGCCGCTGGCCACGT 134
```

```
Db    122 CSGCSTCSCCSCTSACGGCCCAGG 145
          :| :  :| :||:  | ||||||
Qy    135 TCGTGCGGCGCCTGGGGCCCCAGG 158


RESULT  13
LOCUS       A10162        201 bp    DNA             PAT       01-DEC-1993
DEFINITION  Synthetic DNA for preproprotease leader & prochymosin.
ACCESSION   A10162
NID         g489115
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 201)
  AUTHORS   Garvin,R.T. and James,E.
  TITLE     Production of active proteins containing cystine residues
  JOURNAL   Patent: EP 0222279-A 6 20-MAY-1987;
            Cangene Corporation
FEATURES             Location/Qualifiers
     source          1. .201
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
BASE COUNT       22 a     51 c     66 g     16 t     46 others
ORIGIN

  Query Match            0.7%;  Score 30;  DB 17;  Length 201;
  Best Local Similarity  36.1%;  Pred. No. 5.17e-02;
  Matches    52;  Conservative   35;  Mismatches  57;  Indels   0;  Gaps   0;

Db     57 CCTGGGCCGTSAGSGGSGASGCSGASAGSGCSAGGGCCCCSAGSACSGCSGCSCCSCCCT 116
          |||||| |  :||: |:  : |: |:  :||:|| | |  :| : |:  :  :  :|
Cp    158 CCTGGGGCCCCAGGCGCCGCACGAACGTGGCCAGCGGCAGCACCTCGCGGTAGTGGCTGC 99

Db    117 CSAGSAGSCGSCGSGASACSACSGGSCGSGGSCGSCCSGGCTGSGGSGCSGGSGACTCSG 176
          :||:||:  :||:  : |   :  :||:||   :| :|||  :| :| :| :| |  :|
Cp     98 GCAGCAGGGAGCGCACGGCTCGGCAGCGGGGAGCGCGCGGCATCGCGGGGGTGGCCGGGG 39

Db    177 CSGGSGASACSGGSGAGTGSGGCA 200
          |: |:|   |:  :| | |: |||
Cp     38 CCAGGGCTTCCCACGTGCGCAGCA 15


RESULT  14
LOCUS       A10159        201 bp    DNA             PAT       01-DEC-1993
DEFINITION  S.griseus gene for preproprotease (partial).
ACCESSION   A10159
NID         g489114
KEYWORDS    .
SOURCE      Streptomyces griseus.
  ORGANISM  Streptomyces griseus
            Eubacteria; Firmicutes; Actinomycetes; Streptomycetes;
            Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 201)
  AUTHORS   Garvin,R.T. and James,E.
  TITLE     Production of active proteins containing cystine residues
  JOURNAL   Patent: EP 0222279-A 3 20-MAY-1987;
            Cangene Corporation
FEATURES             Location/Qualifiers
     source          1. .201
                     /organism="Streptomyces griseus"
                     /db_xref="taxon:1911"
BASE COUNT       22 a     54 c     66 g     13 t     46 others
ORIGIN

  Query Match            0.7%;  Score 30;  DB 17;  Length 201;
  Best Local Similarity  36.1%;  Pred. No. 5.17e-02;
  Matches    52;  Conservative   35;  Mismatches  57;  Indels   0;  Gaps   0;

Db     57 CCTGGGCCGTSAGSGGSGASGCSGASAGSGCSAGGGCCCCSAGSACSGCSGCSCCSCCCT 116
          |||||| |  :||: |:  : |: |:  :||:|| | |  :| : |:  :  :  :|
Cp    158 CCTGGGGCCCCAGGCGCCGCACGAACGTGGCCAGCGGCAGCACCTCGCGGTAGTGGCTGC 99

Db    117 CSAGSAGSCGSCGSGASACSACSGGSCGSGGSCGSCCSGGCTGSGGSGCSGGSGACTCSG 176
          :||:||:  :||:  : |   :  :||:||   :| :|||  :| :| :| :| |  :|
Cp     98 GCAGCAGGGAGCGCACGGCTCGGCAGCGGGGAGCGCGCGGCATCGCGGGGGTGGCCGGGG 39

Db    177 CSGGSGASACSGGSGAGTGSGGCA 200
          |: |:|   |:  :| | |: |||
Cp     38 CCAGGGCTTCCCACGTGCGCAGCA 15


RESULT  15
LOCUS       E04076        565 bp    RNA             PAT       26-NOV-1996
DEFINITION  gDNA encoding envelope region of type C hepatitis virus.
ACCESSION   E04076
NID         g2172286
KEYWORDS    JP 1992349885-A/1.
SOURCE      Hepatitis C virus.
  ORGANISM  Hepatitis C virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepatitis C-like viruses.
REFERENCE   1  (bases 1 to 565)
  AUTHORS   Tsutae,M., Kazuaki,C., Hiromitsu,K. and Yataro,I. .
  TITLE     NUCLEIC ACID FRAGMENT WITHIN ENVELOPE REGION OF HEPATITIS C VIRUS
            AND METHOD FOR DETECTING THE SAME
  JOURNAL   Patent: JP 1992349885-A 1 04-DEC-1992;
            TEIJIN LTD
COMMENT     OS   Hepatitis C virus
            PN   JP 1992349885-A/1
            PD   04-DEC-1992
            PF   29-MAY-1991 JP 1991152169
            PI   MORINAGA TSUTAE, CHAYAMA KAZUAKI, KUMADA HIROMITSU, PI
            ICHIKAWA YATARO
            PC   C12N15/10,C12Q1/68,C12Q1/70//C12N15/11;
            CC   strandedness: Single;
            CC   topology: Linear;.
FEATURES             Location/Qualifiers
     source          1. .565
                     /organism="Hepatitis C virus"
                     /db_xref="taxon:11103"
BASE COUNT       60 a     93 c    107 g     85 t    220 others
ORIGIN

  Query Match            0.7%;  Score 28;  DB 17;  Length 565;
  Best Local Similarity  22.9%;  Pred. No. 5.96e-01;
  Matches    33;  Conservative   54;  Mismatches  56;  Indels   1;  Gaps   1;

Db    195 GYAYRARGYRRMGGAYRYVATCMTGCWYDYYCCBGGGTGYRYBCCYTGYGTYCGSGARRR 254
          |: :: :  :::   ::::|  : | :::::  :  | |  ::|  ||:   ||:   ::
Cp   1465 GCGCAGGCAGGCCCGCACGAAGCCGTACACCTGCCAGGGGCTGCTGTGCTGGCGGAGCAG 1406

Db    255 YRRYNNYTCBMGNTGYTGGGYRGCGCTBACYCCCACGSTYGCVRCYAGRRA-YVBYAVYV 313
          : ::  :  ::|  |  || : | |  : |: || ||:  ||::|:| ::| :::: : :
Cp   1405 CTGCACCAGGCGACGGGGGTCTGTGTCCTCCTCCTCGGGGGCCGCCACAGAGCCCTGGGG 1346

Db    314 TYCCCRCBRYGVMRMTWCGBCGBC 337
          : |  |:: |:: : :| :||:|
Cp   1345 CTTCTCCCGGGCACAGACACCGGC 1322


Search completed: Mon Aug  3 16:19:16 1998
Job time : 10668 secs.
```

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Mon Aug 3 16:19:35 1998; MasPar time 631.45 Seconds
865.070 Million cell updates/sec

[redacted]bular output not generated.

Title: >US-08-912-951-1
Description: (1-4015) from US08912951.seq
Perfect Score: 4015
N.A. Sequence: 1 GCAGCGCTGCGTCCTGCTGC.....TTTTTCAGTTTTGAAAAAAA 4015
Comp: CGTCGCGACGCAGGACGACG.....AAAAAGTCAAAACTTTTTTT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 188442 seqs, 68026449 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: n-geneseq32
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40

[redacted]istics: Mean 10.074; Variance 6.348; scale 1.587

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	45	1.1	91	9	Q51746	Oligonucleotide probe	2.21e-10
c	2	45	1.1	91	9	Q51746	Oligonucleotide probe	2.21e-10
	3	45	1.1	178	32	T76405	Human endothelin-1 an	2.21e-10
	4	45	1.1	204	1	N81164	Base substituted E.co	2.21e-10
	5	44	1.1	317	32	T76274	Human neutrophil elas	7.74e-10
	6	40	1.0	114	12	Q70466	Generic DNA sequence	1.09e-07
	7	39	1.0	114	12	Q70469	Generic DNA sequence	3.67e-07
	8	41	1.0	172	32	T76363	Human interleukin 8 a	3.20e-08
c	9	40	1.0	172	32	T76363	Human interleukin 8 a	1.09e-07
c	10	41	1.0	204	1	N81164	Base substituted E.co	3.20e-08
	11	39	1.0	264	32	T76445	Substance P receptor	3.67e-07
	12	38	0.9	114	12	Q70465	Generic DNA sequence	1.23e-06
	13	38	0.9	114	12	Q70468	Generic DNA sequence	1.23e-06
	14	36	0.9	114	12	Q70470	Generic DNA sequence	1.33e-05
	15	36	0.9	114	12	Q70467	Generic DNA sequence	1.33e-05
c	16	38	0.9	114	12	Q70465	Generic DNA sequence	1.23e-06
c	17	38	0.9	114	12	Q70467	Generic DNA sequence	1.23e-06
c	18	38	0.9	114	12	Q70468	Generic DNA sequence	1.23e-06
c	19	37	0.9	114	12	Q70469	Generic DNA sequence	4.05e-06
c	20	36	0.9	114	12	Q70466	Generic DNA sequence	1.33e-05
c	21	36	0.9	114	12	Q70470	Generic DNA sequence	1.33e-05
	22	35	0.9	128	32	T76233	Human IL6 antisense o	4.29e-05
	23	35	0.9	162	32	T76307	Human RANTES antisens	4.29e-05
c	24	38	0.9	178	32	T76405	Human endothelin-1 an	1.23e-06
	25	36	0.9	190	32	T76452	Chymase antisense oli	1.33e-05
c	26	35	0.9	190	32	T76452	Chymase antisense oli	4.29e-05
	27	36	0.9	565	6	Q35072	HCV envelope region n	1.33e-05
	28	31	0.8	88	32	T76170	Human IL3 receptor an	4.16e-03
c	29	32	0.8	88	32	T76170	Human IL3 receptor an	1.35e-03
	30	31	0.8	100	32	T76186	Human IL4 receptor an	4.16e-03
	31	34	0.8	114	12	Q70472	Generic DNA sequence	1.37e-04
	32	32	0.8	114	12	Q70471	Generic DNA sequence	1.35e-03
	33	31	0.8	114	12	Q70473	Generic DNA sequence	4.16e-03
c	34	32	0.8	114	12	Q70472	Generic DNA sequence	1.35e-03
c	35	31	0.8	114	12	Q70473	Generic DNA sequence	4.16e-03
c	36	31	0.8	128	32	T76233	Human IL6 antisense o	4.16e-03
c	37	32	0.8	145	32	T76299	Human defensin 3 anti	1.35e-03
	38	34	0.8	168	32	T76270	Human MDNCF antisense	1.37e-04
	39	34	0.8	172	32	T76376	Human GM-CSF antisens	1.37e-04
c	40	30	0.7	100	32	T76186	Human IL4 receptor an	1.26e-02
c	41	30	0.7	114	12	Q70471	Generic DNA sequence	1.26e-02
c	42	30	0.7	130	32	T76152	Human vascular cell a	1.26e-02
c	43	30	0.7	168	32	T76270	Human MDNCF antisense	1.26e-02
	44	30	0.7	201	2	N70195	Streptomyces protease	1.26e-02
	45	30	0.7	201	2	N70194	Signal portion of gen	1.26e-02

ALIGNMENTS

RESULT 1
ID Q51746 standard; cDNA; 91 BP.
AC Q51746;
DT 31-MAY-1994 (first entry)
DE Oligonucleotide probe MK14-A
KW Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW ss.
OS Synthetic.
PN EP-571911-A.
PD 01-DEC-1993.
PF 24-MAY-1993; 108325.
PR 26-MAY-1992; US-889651.
PA (BECT) BECTON DICKINSON CO.
PI Shank DD, Spears PA;
DR WPI; 93-378844/48.
PT New oligo:nucleotide probes specific for Mycobacteria - used for
PT detection and amplification of Mycobacteria nucleic acid in
PT samples
PS Claim 3; Page 14; 23pp; English.
CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC (Q51735). It hybridized to all spp. of mycobacteria tested, but
CC cross reacted to a few non-mycobacterial spp. The probe may
CC be useful as an initial screen for mycobacterial infection.
CC See also Q51735-45 and Q51747-59.
SQ Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;

Query Match 1.1%; Score 45; DB 9; Length 91;
Best Local Similarity 7.3%; Pred. No. 2.21e-10;
Matches 4; Conservative 46; Mismatches 5; Indels 0; Gaps 0;

Db 6 cggcgssvhsyyvvhvvshhhsvhhvvhhvhvsvvvvhhvvhvvhhvhyhvyvsv 60
||| |::::::::::::::::::::::::::: ::::: : :::: :::::::::
Qy 379 CGGGGGCCCCCCCGAGGCCTTCACCACCAGCGTGCGCAGCTACCTGCCCAACACG 433

RESULT 2
ID Q51746 standard; cDNA; 91 BP.

```
AC   Q51746;
DT   31-MAY-1994  (first entry)
DE   Oligonucleotide probe MK14-A
KW   Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW   ss.
OS   Synthetic.
PN   EP-571911-A.
PD   01-DEC-1993.
PF   24-MAY-1993; 108325.
PR   26-MAY-1992; US-889651.
PA   (BECT ) BECTON DICKINSON CO.
PI   Shank DD,  Spears PA;
DR   WPI; 93-378844/48.
PT   New oligo:nucleotide probes specific for Mycobacteria - used for
PT   detection and amplification of Mycobacteria nucleic acid in
PT   samples
PS   Claim 3; Page 14; 23pp; English.
CC   Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC   (Q51735).  It hybridized to all spp. of mycobacteria tested, but
CC   cross reacted to a few non-mycobacterial spp.  The probe may
CC   be useful as an initial screen for mycobacterial infection.
CC   See also Q51735-45 and Q51747-59.
SQ   Sequence   91 BP;    5 A;    17 C;    15 G;    4 T;

  Query Match             1.1%;  Score 45;  DB 9;  Length 91;
  Best Local Similarity  11.9%;  Pred. No. 2.21e-10;
  Matches    7;  Conservative   45;  Mismatches   7;  Indels   0;  Gaps   0;

Db      4 tccggcgssvhsyyvvhvvshhhsvhhvvhhvhvsvvvvhhvvhvvhhvhyhvyvsvct 62
          || ||| ::::::::::::::::  :::::::: :::::: :::: ::::::::::::|| 
Cp   1118 TCAGGCTGGGCCTCAGAGAGCTGAGTAGGAAGGAGGGCCGCAGCTGCTCCTTGTCGCCT 1060


RESULT    3
ID   T76405 standard; DNA; 178 BP.
AC   T76405;
DT   15-SEP-1997  (first entry)
DE   Human endothelin-1 antisense oligonucleotide.
KW   Asthma; airway epithelium; adenosine free; cystic fibrosis;
KW   chronic obstructive pulmonary disease; bronchitis; ss.
OS   Synthetic.
PN   WO9640162-A1.
PD   19-DEC-1996.
PF   06-JUN-1996; U09306.
PR   07-JUN-1995; US-474497.
PA   (UYEC-) UNIV EAST CAROLINA.
PI   Metzger WJ,  Nyce JW;
DR   WPI; 97-051871/05.
PT   Treatment of airway diseases such as asthma - by topically applying
PT   adenosine-free antisense oligo:nucleotide to airway epithelium of
PT   subject
PS   Claim 5; Page 38; 71pp; English.
CC   A method for treating airway disease in a subject has been produced,
CC   which involves the topical administration of an essentially adenosine
CC   free antisense oligonucleotide (ON) to the airway epithelium of the
CC   subject. The present sequence is an antisense oligonucleotide specific
CC   for the human endothelin-1, targeted at the initiation codon. The
CC   method can be used to treat airway diseases such as cystic fibrosis,
CC   asthma, chronic obstructive pulmonary disease, bronchitis and other
CC   airway diseases characterised by an inflammatory response. By
CC   eliminating adenosine from the antisense ON, its liberation upon
CC   antisense degradation is prevented, thereby preventing adenosine-
CC   induced bronchoconstriction in patients with hyper-reactive airways.
SQ   Sequence   178 BP;    0 A;    52 C;    46 G;    32 T;

  Query Match             1.1%;  Score 45;  DB 32;  Length 178;
  Best Local Similarity  37.3%;  Pred. No. 2.21e-10;
  Matches   44;  Conservative   40;  Mismatches  33;  Indels   1;  Gaps   1;

Db     53 bcccbccgcgctgbgctcbgcgcctbbgbctgctgtttctggbgctccttggcbbgccbc 112
          :|| :|| |||  :||  :|||| :: :|||| |   |  |:|||| ||  ::|| :| 
Qy     44 GCCACCCCCGCGATGCCGCGCGCTCCCCGCTGCCGAGCCGTGCGCTCCCTGCTGCGCAGC 103

Db    113 bbbcbgcbgbgbgbbbbtcbtgbgcbbbtbbtccbttctgbbbbbbbgggbtcbbbbb 170
          : : : |:|:| |::::| : |:  ::: ::| |:| | |::::::::|| : | ::::
Qy    104 CAC-TACCGCGAGGTGCTGCCGCTGGCCACGTTCGTGCGGCGCCTGGGGCCCCAGGGC 160


RESULT    4
ID   N81164 standard; DNA; 204 BP.
AC   N81164;
DT   08-NOV-1990  (first entry)
DE   Base substituted E.coli beta-galactosidase alpha-fragment.
KW   E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS   Escherichia coli.
FH   Key             Location/Qualifiers
FT   misc_feature    19..69
FT                   /*tag=  a
FT                   /function=multiple cloning site
FT   primer_bind     187..204
FT                   /*tag=  b
PN   EP-285123-A.
PD   05-MAY-1988.
PF   30-MAR-1988; 105163.
PR   03-APR-1987; US-034819.
PA   (SUSO) SUOMEN SOKERI OY.
PI   Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
DR   WPI; 88-279927/40.
PT   Introducing random point mutations into nucleic acods -
PT   by prepn of single stranded template, annealing a primer, elongation,
PT   misincorporation, completion of molecules and screening.
PS   Disclosure; p; English.
CC   Random point mutations were introduced into the alpha fragment of
CC   E.coli beta-galactosidase. The wild type sequence was obtained as a
CC   single stranded template and an oligonucleotide was hybridised to
CC   it to generate a popn of DNA molecules which terminate at all
CC   possible nucleotide positions within a specified region. The
CC   variable 3' ends generated in this way are used as primers for
CC   reverse transcriptase. Nucleotides are misincorporated by the
CC   transcriptase and the molecules are completed to forms that can be
CC   amplified and then expressed in a suitable host-vector system.
CC   The sequence covers all 176 difft base substitutions, most of which
CC   occurred singularly in any given mutant.
CC   See also P80575.
SQ   Sequence   204 BP;   21 A;   47 C;   17 G;   11 T;   108 Others;

  Query Match             1.1%;  Score 45;  DB 1;  Length 204;
  Best Local Similarity  12.5%;  Pred. No. 2.21e-10;
  Matches   12;  Conservative   54;  Mismatches  29;  Indels   1;  Gaps   1;

Db     91 thhyrrmrbnvyrdynrsdaaawyccyrrsvkydccynachhddh-yvybbbvynvhnhn 149
          |:::: ::: : ::: :::| |::||::  ::: |:  | :::: :::::::: :  :
Qy   3116 TTTCATCAGCAAGTTTGGAAGAACCCCACATTTTTCCTGCGCGTCATCTCTGACACGGCC 3175

Db    150 ncncccbnnhvchnvhbnnhrnwayvrhdarrddvh 185
           | | |:  ::|: :::  :: :| : :: :: :::
Qy   3176 TCCCTCTGCTACTCCATCCTGAAAGCCAAGAACGCA 3211


RESULT    5
ID   T76274 standard; DNA; 317 BP.
AC   T76274;
DT   15-SEP-1997  (first entry)
DE   Human neutrophil elastase (medullasin) antisense oligonucleotide.
KW   Asthma; airway epithelium; adenosine free; cystic fibrosis;
KW   chronic obstructive pulmonary disease; bronchitis; ss.
OS   Synthetic.
PN   WO9640162-A1.
PD   19-DEC-1996.
PF   06-JUN-1996; U09306.
PR   07-JUN-1995; US-474497.
PA   (UYEC-) UNIV EAST CAROLINA.
PI   Metzger WJ,  Nyce JW;
DR   WPI; 97-051871/05.
PT   Treatment of airway diseases such as asthma - by topically applying
PT   adenosine-free antisense oligo:nucleotide to airway epithelium of
```

```
PT   subject
PS   Claim 5; Page 33; 71pp; English.
CC   A method for treating airway disease in a subject has been produced,
CC   which involves the topical administration of an essentially adenosine
CC   free antisense oligonucleotide (ON) to the airway epithelium of the
CC   subject. The present sequence is an antisense oligonucleotide specific
CC   for the human neutrophil elastase (medullasin), targeted at the
CC   initiation codon. The method can be used to treat airway diseases
CC   such as cystic fibrosis, asthma, chronic obstructive pulmonary disease,
CC   bronchitis and other airway diseases characterised by an inflammatory
CC   response. By eliminating adenosine from the antisense ON, its liberation
CC   upon antisense degradation is prevented, thereby preventing adenosine-
CC   induced bronchoconstriction in patients with hyper-reactive airways.
SQ   Sequence   317 BP;    0 A;    112 C;    118 G;    36 T;

  Query Match             1.1%;  Score 44;  DB 32;  Length 317;
  Best Local Similarity  46.4%;  Pred. No. 7.74e-10;
  Matches   122;  Conservative   44;  Mismatches  92;  Indels   5;  Gaps   5;

Db     35 bccbcccgcbccgcgcggbcgtttbcbttcgccbcgcbgtgcgcggccgbcbtgbcgbbg 94
           :||:| |||:|| ||| |:||   :|:| ||| :|  : ||||| ||| :    :|| :|
Qy     58 GCCGCGCGCTCCCCGCTGCCGAG-CCGTGCGCTCCCTGCTGCGCAGCCACTACCGCGAGG 116

Db     95 ttgggc-gcbbtcbgggtg-gcgccgcbgbbgtggcctccgcgcbgctgcbgggbcbccb 152
          |   || ||:: |:  ||  | || ||: ::| |||| | | ||:|  | : ||: :  :
Qy    117 TGCTGCCGCTGGCCACGTTCGTGCGGCGCCTGGGGCCCCAGGGCTGGCGGCTGGTGCAGC 176

Db    153 tgbbgggccbcgcgtggggccgcgctc-gccggccccccbcbbtctccgbggccbgcgcg 211
            ::|| ||: |||     ||||| | |  |||||   :|::  |  |:|  |:  | |
Qy    177 GCGGGGACCCGGCGGCTTTCCGCGCGCTGGTGGCCCAGTGCCTGGTGTGCGTGCCCTGGG 236

Db    212 gtgccccccbgcbgcbbggccggcbggbcbcbggcgbggbgbcbcgcgbgtcggcggccg 271
            || |  |:||: |::| |   |:  :|: : ||  ||:|:|: || :|  || | | |
Qy    237 ACGCACGGCCGCCCCCCGCCGCCCCCTCCTTCCGCCAGGTGTCCTGCCTGAAGGAG-CTG 295

Db    272 bgggtcbtggtggggctggggct 294
          : || |:  |||  || | ||||
Qy    296 GTGGCCCGAGTGCTGCAGAGGCT 318


RESULT    6
ID   Q70466 standard; DNA; 114 BP.
AC   Q70466;
DT   05-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR-9 petide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag=  a
FT                   /note= "this sequence represents 'Z'; Z can be a
FT                   sequence of 6, 9 or 12 nucleotides (see
FT                   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R65152.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70466 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented as follows:X(NNB)1(TGC)(NNB)10(TGC)2(NNB)4Z(NNB)8(TGC)(NNB)
CC   -9Y. X and Y are flanking  restriction sites (X is not the same as Y)
CC   that are not specified further. Other generic sequences are shown in
CC   Q70466-68. Other specific peptides generated by these generic sequences
CC   are shown in R65151-54. TSARs are concatenated heterofunctional proteins
CC   or peptides, comprising at least two functional regions - a binding
CC   domain with affinity for a ligand and a second effector peptide portion
CC   that is chemically or biologically active.They may further comprise a
CC   linker peptide between the 2 domains.The oligonucleotides are also
CC   designed so that the expressed peptide contains 2 or 4 cysteine residues
CC   positioned in, or flanking, the unpredicted or variant residues. These
CC   residues confer some degree of conformational rigidity to the peptides.
CC   The TSARs or compsns. comprising a TSAR binding domain can be used in
CC   vivo to deliver a chemically or biologically active moiety, eg. metal
CC   ion, radioisotope, peptide, toxin or enzyme, to the specific target or
CC   on the cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need for
CC   complex methods of hybridoma formation or in vivo antibody production.
CC   The TSARs are easily characterised and have designed activity allowing
CC   direct and rapid detection in a screening process.
SQ   Sequence   114 BP;    0 A;    4 C;    4 G;    4 T;

  Query Match             1.0%;  Score 40;  DB 12;  Length 114;
  Best Local Similarity   8.9%;  Pred. No. 1.09e-07;
  Matches    10;  Conservative   32;  Mismatches  70;  Indels   0;  Gaps   0;

Db      3 btgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgctgcnnbnnbnnbnnbnnnnnnnn 62
            :|||  :  :  :  :  :  :  :  :  :  :  :||| ||  :  :  :  :
Qy    137 GTGCGGCGCCTGGGGCCCCAGGGCTGGCGGCTGGTGCAGCGCGGGGACCCGGCGGCTTTC 196

Db     63 bnnbnnbnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnbnnbnnb 114
          :  :  :  :  :  :  :  :||  :  :  :  :  :  :  :  :  :
Qy    197 CGCGCGCTGGTGGCCCAGTGCCTGGTGTGCGTGCCCTGGGACGCACGGCCGC 248


RESULT    7
ID   Q70469 standard; DNA; 114 BP.
AC   Q70469;
DT   07-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR peptide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag=  a
FT                   /note= "this sequence represents 'Z'; Z can be a
FT                   sequence of 6,9 or 12 nucleotides (see
FT                   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70469 is a generic DNA sequence used to generate random TSAR peptide
CC   This generic formula can be represented  as follows: X(TGC)(NNB)10-
CC   (TGC)(NNB)6Z(NNB)2(TGC)(NNB)14(TGC)Y. X and Y are flanking  restriction
CC   sites (X is not the same as Y) that are not specified further. This
CC   sequence generates peptides that are cloverleaf in structure. Other
CC   generic sequences are shown in Q70465-68. Other specific peptides
CC   generated by these generic sequences are shown in R65150-54. TSARs are
CC   concatenated heterofunctional proteins or peptides, comprising at least
CC   two functional regions - a binding domain with affinity for a ligand and
CC   a second effector peptide portion that is chemically or biologically
CC   active.They may further comprise a linker peptide between the 2 domains.
CC   The oligonucleotides are also designed so that the expressed peptide
```

```
CC   contains 2 or 4 cysteine residues positioned in, or flanking, the
CC   unpredicted or variant residues. These residues confer some degree of
CC   conformational rigidity to the peptides. The TSARs or compsns. comprising
CC   a TSAR binding domain can be used in vivo to deliver a chemically or
CC   biologically active moiety, eg. metal ion, radioisotope, peptide, toxin
CC   or enzyme, to the specific target or on the cell. They can also replace
CC   the function of macromolecules, eg. monoclonal or polyclonal antibodies
CC   and therefore circumvent the need for complex methods of hybridoma
CC   formation or in vivo antibody production. The TSARs are easily
CC   characterised and have designed activity allowing direct and rapid
CC   detection in a screening process.
SQ   Sequence   114 BP;    0 A;    4 C;    4 G;    4 T;

  Query Match            1.0%;  Score 39;  DB 12;  Length 114;
  Best Local Similarity  7.2%;  Pred. No. 3.67e-07;
  Matches     8;  Conservative   32;  Mismatches  71;  Indels   0;  Gaps   0;

Db      1 tgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnnnn 60
          |||  :  :  :  :  :  :  :  :  :|||  :  :  :  :  :  :  :
Qy    138 TGCGGCGCCTGGGGCCCCAGGGCTGGCGGCTGGTGCAGCGCGGGGACCCGGCGGCTTTCC 197

Db     61 nnbnnbtgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 111
           :  :||   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Qy    198 GCGCGCTGGTGGCCCAGTGCCTGGTGTGCGTGCCCTGGGACGCACGGCCGC 248


RESULT    8
ID   T76363 standard; DNA; 172 BP.
AC   T76363;
DT   15-SEP-1997  (first entry)
DE   Human interleukin 8 antisense oligonucleotide.
KW   Asthma; airway epithelium; adenosine free; cystic fibrosis;
KW   chronic obstructive pulmonary disease; bronchitis; ss.
OS   Synthetic.
PN   WO9640162-A1.
PD   19-DEC-1996.
PF   06-JUN-1996; U09306.
PR   07-JUN-1995; US-474497.
PA   (UYEC-) UNIV EAST CAROLINA.
PI   Metzger WJ,  Nyce JW;
DR   WPI; 97-051871/05.
PT   Treatment of airway diseases such as asthma - by topically applying
PT   adenosine-free antisense oligo:nucleotide to airway epithelium of
PT   subject
PS   Claim 5; Page 36; 71pp; English.
CC   A method for treating airway disease in a subject has been produced,
CC   which involves the topical administration of an essentially adenosine
CC   free antisense oligonucleotide (ON) to the airway epithelium of the
CC   subject. The present sequence is an antisense oligonucleotide specific
CC   for the human interleukin 8, targeted at the initiation codon. The
CC   method can be used to treat airway diseases such as cystic fibrosis,
CC   asthma, chronic obstructive pulmonary disease, bronchitis and other
CC   airway diseases characterised by an inflammatory response. By
CC   eliminating adenosine from the antisense ON, its liberation upon
CC   antisense degradation is prevented, thereby preventing adenosine-
CC   induced bronchoconstriction in patients with hyper-reactive airways.
SQ   Sequence   172 BP;    0 A;    35 C;    42 G;    39 T;

  Query Match            1.0%;  Score 41;  DB 32;  Length 172;
  Best Local Similarity  32.2%;  Pred. No. 3.20e-08;
  Matches    47;  Conservative   49;  Mismatches  49;  Indels   1;  Gaps   1;

Db     22 bbgbbtbgctttgctbtctbbggbtcbcbtttbgbcbtbggbbbbcgctgtbggtcbgbb 81
          :: :: :|||  | :  |:: |:  : : | :|:|: : |:::: || | :||| : ::
Qy    117 TGCTGCCGCTGGCCACGTTCGTGCGGCGCCTGGGGCCCCAGGGCTGGCGGCTGGTGCAGC 176

Db     82 bgbtgtgcttbccttcbcbcbgbgctgcbgbbbtcbggbbggctgccbbgbgbgccbcgg 141
          : : |  |  : |  |: :|:|:|| ||:|::: |:  ::| |||  ::|:|:|||: ||
Qy    177 GCGGGGACCCGGCGGCTTTCCGCGC-GCTGGTGGCCCAGTGCCTGGTGTGCGTGCCCTGG 235

Db    142 ccbgcttggbgtcbtgtttbcbcbcb 167
            :||  ||:  |:     :|:|:|:
Qy    236 GACGCACGGCCGCCCCCCGCCGCCCC 261
```

```
RESULT    9
ID   T76363 standard; DNA; 172 BP.
AC   T76363;
DT   15-SEP-1997  (first entry)
DE   Human interleukin 8 antisense oligonucleotide.
KW   Asthma; airway epithelium; adenosine free; cystic fibrosis;
KW   chronic obstructive pulmonary disease; bronchitis; ss.
OS   Synthetic.
PN   WO9640162-A1.
PD   19-DEC-1996.
PF   06-JUN-1996; U09306.
PR   07-JUN-1995; US-474497.
PA   (UYEC-) UNIV EAST CAROLINA.
PI   Metzger WJ,  Nyce JW;
DR   WPI; 97-051871/05.
PT   Treatment of airway diseases such as asthma - by topically applying
PT   adenosine-free antisense oligo:nucleotide to airway epithelium of
PT   subject
PS   Claim 5; Page 36; 71pp; English.
CC   A method for treating airway disease in a subject has been produced,
CC   which involves the topical administration of an essentially adenosine
CC   free antisense oligonucleotide (ON) to the airway epithelium of the
CC   subject. The present sequence is an antisense oligonucleotide specific
CC   for the human interleukin 8, targeted at the initiation codon. The
CC   method can be used to treat airway diseases such as cystic fibrosis,
CC   asthma, chronic obstructive pulmonary disease, bronchitis and other
CC   airway diseases characterised by an inflammatory response. By
CC   eliminating adenosine from the antisense ON, its liberation upon
CC   antisense degradation is prevented, thereby preventing adenosine-
CC   induced bronchoconstriction in patients with hyper-reactive airways.
SQ   Sequence   172 BP;    0 A;    35 C;    42 G;    39 T;

  Query Match            1.0%;  Score 40;  DB 32;  Length 172;
  Best Local Similarity  29.7%;  Pred. No. 1.09e-07;
  Matches    47;  Conservative   52;  Mismatches  59;  Indels   0;  Gaps   0;

Db     11 bccbbbgcbtcbbgbbtbgctttgctbtctbbggbtcbcbtttbgbcbtbggbbbbcgct 70
          :||::: |: |::|:: :|    || :   ::| :  : :|  :|: :|: |::::||
Cp   3535 CCCCTCCCTCCCTGGGACGTAGAGCCCGGCGTGACAGGGCTGCTGGTGTCTGCTCTCGGC 3476

Db     71 gtbggtcbgbbbgbtgtgcttbccttcbcbcbgbgctgcbgbbbtcbggbbggctgccbb 130
           |:| | :|:::|:   ||  :|  ||: :   :|    :  ::  : |::   ||||::
Cp   3475 CTGGCTGTGGGCGGGTGGCCATCAGTCCAGGATGGTCTTGAAGTCTGAGGGCAGTGCCGG 3416

Db    131 gbgbgccbcggccbgcttggbgtcbtgtttbcbcbcbg 168
          |: :|| :|||||: |  ||:   : |  |: :|:|:|
Cp   3415 GTTGGCTGCGGCCTCCAGGGCAGTCAGCGTCGTCCCCG 3378


RESULT   10
ID   N81164 standard; DNA; 204 BP.
AC   N81164;
DT   08-NOV-1990  (first entry)
DE   Base substituted E.coli beta-galactosidase alpha-fragment.
KW   E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS   Escherichia coli.
FH   Key             Location/Qualifiers
FT   misc_feature    19..69
FT                   /*tag=  a
FT                   /function=multiple cloning site
FT   primer_bind     187..204
FT                   /*tag=  b
PN   EP-285123-A.
PD   05-MAY-1988.
PF   30-MAR-1988; 105163.
PR   03-APR-1987; US-034819.
PA   (SUSO) SUOMEN SOKERI OY.
PI   Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
DR   WPI; 88-279927/40.
PT   Introducing random point mutations into nucleic acods -
PT   by prepn of single stranded template, annealing a primer, elongation,
```

```
PT   misincorporation, completion of molecules and screening.
PS   Disclosure; p; English.
CC   Random point mutations were introduced into the alpha fragment of
CC   E.coli beta-galactosidase. The wild type sequence was obtained as a
CC   single stranded template and an oligonucleotide was hybridised to
CC   it to generate a popn of DNA molecules which terminate at all
CC   possible nucleotide positions within a specified region. The
CC   variable 3' ends generated in this way are used as primers for
CC   reverse transcriptase. Nucleotides are misincorporated by the
CC   transcriptase and the molecules are completed to forms that can be
CC   amplified and then expressed in a suitable host-vector system.
CC   The sequence covers all 176 difft base substitutions, most of which
CC   occurred singularly in any given mutant.
CC   See also P80575.
SQ   Sequence   204 BP;   21 A;   47 C;   17 G;   11 T;   108 Others;

  Query Match            1.0%;  Score 41;  DB 1;  Length 204;
  Best Local Similarity  18.8%;  Pred. No. 3.20e-08;
  Matches    27;  Conservative   58;  Mismatches  59;  Indels   0;  Gaps   0;

Db     50 aggnyccccgggywccgagcycgaayycdchvgccgymrttthhyrrmrbnvyrdynrsd 109
          |||   || |||: ||| ||:||  :  : : |    :: | ::::: ::  : ::
Cp   2082 AGGAGGCCGGGGCGCCGCGCCCGCTCGTAGTTGAGCACGCTGAACAGTGCCTTCACCCTC 2023

Db    110 aaawyccyrrsvkydccynachhddhyvybbbvynvhnhnncncccbnnhvchnvhbnnh 169
          |  :   : ::::: |:  |:::::::::::  :  :    | |:  :: :  ::  :
Cp   2022 GAGGTGAGACGCTCGGCCCTCTTTTCTCTGCGGAACGTTCTGGCTCCCACGACGTAGTCC 1963

Db    170 rnwayvrhdarrddvhccvchccg 193
          :   :::::|  :::: |: :|||
Cp   1962 ATGTTCACAATCGGCCGCAGCCCG 1939


RESULT   11
ID   T76445 standard; DNA; 264 BP.
AC   T76445;
DT   16-SEP-1997  (first entry)
DE   Substance P receptor antisense oligonucleotide.
KW   Asthma; airway epithelium; adenosine free; cystic fibrosis;
KW   chronic obstructive pulmonary disease; bronchitis; ss.
OS   Synthetic.
PN   WO9640162-A1.
PD   19-DEC-1996.
PF   06-JUN-1996; U09306.
PR   07-JUN-1995; US-474497.
PA   (UYEC-) UNIV EAST CAROLINA.
PI   Metzger WJ,  Nyce JW;
DR   WPI; 97-051871/05.
PT   Treatment of airway diseases such as asthma - by topically applying
PT   adenosine-free antisense oligo:nucleotide to airway epithelium of
PT   subject
PS   Example 5; Page 40; 71pp; English.
CC   A method for treating airway disease in a subject has been produced,
CC   which involves the topical administration of an essentially adenosine
CC   free antisense oligonucleotide (ON) to the airway epithelium of the
CC   subject. The present sequence is an antisense oligonucleotide specific
CC   for the substance P receptor, targeted at the initiation codon. The
CC   method can be used to treat airway diseases such as cystic fibrosis,
CC   asthma, chronic obstructive pulmonary disease, bronchitis and other
CC   airway diseases characterised by an inflammatory response. By
CC   eliminating adenosine from the antisense ON, its liberation upon
CC   antisense degradation is prevented, thereby preventing adenosine-induced
CC   bronchoconstriction in patients with hyper-reactive airways.
SQ   Sequence   264 BP;    0 A;    72 C;    77 G;    56 T;

  Query Match            1.0%;  Score 39;  DB 32;  Length 264;
  Best Local Similarity  37.9%;  Pred. No. 3.67e-07;
  Matches    83;  Conservative   51;  Mismatches  83;  Indels   2;  Gaps   2;

Db     13 ccbcbtcbctbccbcgttgcccbccbcbgbggtcbccbcbbtgbccgtgtbggcbgctgc 72
          ||: : |:| : |:     | |:  :|:|:|  |:||:|::||:| | | :|  :||| |
Qy     33 CCCTGGCCCCGGCCACCCCCGCGATGCCGCGCGCTCCCCGCTGCC-GAGCCGTGCGCTCC 91

Db     73 ccbbbggbcbbtttgccbggctggttgcbcgbbctgbttgggttccgbggtgttbgtggb 132
          | :::| :| :      : ||  | |||:  ::|||:    |||| |:|  |  : |||:
Qy     92 CTGCTGCGCAGCCACTACCGCGAGGTGCTGCCGCTGGCCACGTTC-GTGCGGCGCCTGGG 150

Db    133 gbtgtttggggbgbggtctgbgtccbccgggbggbcgttbtccbtttcgbbgctbggcgg 192
          |:     ||  :|:||   |:|   : ||||:  :||  : | :|    ::|||:|  |
Qy    151 GCCCCAGGGCTGGCGGCTGGTGCAGCGCGGGGACCCGGCGGCTTTCCGCGCGCTGGTGGC 210

Db    193 tbbbgccctbctbtctgtbcbcbbccccctctgcbgcb 231
          : :  |||: |:|  ||:|:|::   | | | ||:||:
Qy    211 CCAGTGCCTGGTGTGCGTGCCCTGGGACGCACGGCCGCC 249


RESULT   12
ID   Q70465 standard; DNA; 114 BP.
AC   Q70465;
DT   05-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR petide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag= a
FT                   /note= "this sequence represents 'Z'; Z can be a
FT                   sequence of 6, 9 or 12 nucleotides (see
FT                   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R65150 and R65151.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70465 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)6(TGC)(NNB)11Z(NNB)14(TGC)(NNB)3Y. X
CC   and Y are flanking  restriction sites (X is not the same as Y) that are
CC   not specified further. Other generic sequences are shown in Q70466-68.
CC   Other specific peptides generated by these generic sequences are shown in
CC   R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC   comprising at least two functional regions - a binding domain with
CC   affinity for a ligand and a second effector peptide portion that is
CC   chemically or biologically active.They may further comprise a linker
CC   peptide between the 2 domains.The oligonucleotides are also designed so
CC   that the expressed peptide contains 2 or 4 cysteine residues positioned
CC   in, or flanking, the unpredicted or variant residues. These residues
CC   confer some degree of conformational rigidity to the peptides. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need
CC   for complex methods of hybridoma formation or in vivo antibody
CC   production. The TSARs are easily characterised and have designed
CC   activity allowing direct and rapid detection in a screening process.
SQ   Sequence   114 BP;    0 A;    2 C;    2 G;    2 T;

  Query Match            0.9%;  Score 38;  DB 12;  Length 114;
  Best Local Similarity  5.4%;  Pred. No. 1.23e-06;
  Matches     6;  Conservative   33;  Mismatches  73;  Indels   0;  Gaps   0;

Db      3 bnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnnnnnn 62
          :  :     :  :  :|||  :  :  :  :  :  :  :  :  :  :  :
Qy   2479 CTTCCTACGCTTCATGTGCCACCACGCCGTGCGCATCAGGGGCAAGTCCTACGTCCAGTG 2538
```

```
Db      63 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnbnnb 114
           :  :  :  :  :  :  :  :  :  :  :  :  :  :  :|||  :  :  :
Qy    2539 CCAGGGGATCCCGCAGGGCTCCATCCTCTCCACGCTGCTCTGCAGCCTGTGC 2590


RESULT   13
ID   Q70468 standard; DNA; 114 BP.
AC   Q70468;
DT   05-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR petide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concatenated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag=  a
FT                   /note= "this sequence represents 'Z'; Z can be a
                     sequence of 6, 9 or 12 nucleotides (see
                     comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R65154.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70468 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)11(TGC)(NNB)6Z(NNB)7(TGC)(NNB)10Y. X
CC   and Y are flanking  restriction sites (X is not the same as Y) that are
CC   not specified further. Other generic sequences are shown in Q70466-68.
CC   Other specific peptides generated by these generic sequences are shown in
CC   R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC   comprising at least two functional regions - a binding domain with
CC   affinity for a ligand and a second effector peptide portion that is
CC   chemically or biologically active.They may further comprise a linker
CC   peptide between the 2 domains.The oligonucleotides are also designed so
CC   that the expressed peptide contains 2 or 4 cysteine residues positioned
CC   in, or flanking, the unpredicted or variant residues. These residues
CC   confer some degree of conformational rigidity to the peptides. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need
CC   for complex methods of hybridoma formation or in vivo antibody
CC   production. The TSARs are easily characterised and have designed activity
CC   allowing direct and rapid detection in a screening process.
SQ   Sequence   114 BP;    0 A;    2 C;    2 G;    2 T;

  Query Match             0.9%;  Score 38;  DB 12;  Length 114;
  Best Local Similarity   4.5%;  Pred. No. 1.23e-06;
  Matches     5;  Conservative   34;  Mismatches  73;  Indels   0;  Gaps   0;

Db       3 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnnnnnnn 62
           :  :  :  :  :  :  :  :  :  :  :  :|||  :  :  :  :  :  :  :
Qy     140 CGGCGCCTGGGGCCCCAGGGCTGGCGGCTGGTGCAGCGCGGGGACCCGGCGGCTTTCCGC 199

Db      63 bnnbnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 114
           :  :  :  :  :  :  :||   :  :  :  :  :  :  :  :  :  :
Qy     200 GCGCTGGTGGCCCAGTGCCTGGTGTGCGTGCCCTGGGACGCACGGCCGCCCC 251


RESULT   14
ID   Q70470 standard; DNA; 114 BP.
AC   Q70470;
DT   10-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR peptide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concatenated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag=  a
FT                   /note= "encoded by Z (see comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R58378.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 36; 255pp; English.
CC   Q70470 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)4(CAC)(NNB)4(CAC)(NNB)8Z(NNB)6(CAC)(NNB)8
CC   -(CAC)2(NNB)Y. X and Y are flanking  restriction sites (X is not the same
CC   as Y) that are not specified further. The peptides generated by this and
CC   other generic sequences (Q70471-73) have invariant histidine residues
CC   incorporated into variant sequences. TSARs are concatenated
CC   heterofunctional proteins or peptides, comprising at least two functional
CC   regions - a binding domain with affinity for a ligand and a second
CC   effector peptide portion that is chemically or biologically active.They
CC   may further comprise a linker peptide between the 2 domains. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need
CC   for complex methods of hybridoma formation or in vivo antibody
CC   production. The TSARs are easily characterised and have designed
CC   activity allowing direct and rapid detection in a screening process.
SQ   Sequence   114 BP;    5 A;    10 C;    0 G;    0 T;

  Query Match             0.9%;  Score 36;  DB 12;  Length 114;
  Best Local Similarity   9.8%;  Pred. No. 1.33e-05;
  Matches    11;  Conservative   30;  Mismatches  71;  Indels   0;  Gaps   0;

Db       3 bnnbnnbnnbcacnnbnnbnnbnnbcacnnbnnbnnbnnbnnbnnbnnbnnbnnnnnnnn 62
           :  :  :  : ||  :  :  :  :|||  :  :     :  :  :  :  :
Qy     481 CCGCGTGGGCGACGACGTGCTGGTTCACCTGCTGGCACGCTGCGCGCTCTTTGTGCTGGT 540

Db      63 bnnbnnbnnbnnbnnbcacnnbnnbnnbnnbnnbnnbnnbnnbcaccacnnb 114
           :  :  :  :  :  : ||  :  :  :  :  :  :  :  :|| | |  :
Qy     541 GGCTCCCAGCTGCGCCTACCAGGTGTGCGGGCCGCCGCTGTACCAGCTCGGC 592


RESULT   15
ID   Q70467 standard; DNA; 114 BP.
AC   Q70467;
DT   05-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR petide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concatenated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag=  a
FT                   /note= "this sequence represents 'Z'; Z can be a
```

FT sequence of 6, 9 or 12 nucleotides (see
FT comments)"
PN WO9418318-A.
PD 18-AUG-1994.
PF 01-FEB-1994; U00977.
PR 01-FEB-1993; US-013416.
PR 30-DEC-1993; US-176500.
PR 31-JAN-1994; US-189331.
PA (UYNC-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
DR WPI; 94-279739/34.
DR P-PSDB; R65153.
PT Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PT comprising a binding domain and an effector domain
PS Disclosure; Page 35; 255pp; English.
CC Q70467 is a generic DNA sequence used to generate random TSAR (Totally
CC Synthetic Affinity Reagents) peptides.This generic formula can also be
CC represented as follows: X(NNB)16(TGC)(NNB)1Z(NNB)16(TGC)(NNB)1Y. X
CC and Y are flanking restriction sites (X is not the same as Y) that are
CC not specified further. Other generic sequences are shown in Q70466-68.
CC Other specific peptides generated by these generic sequences are shown in
CC R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC comprising at least two functional regions - a binding domain with
CC affinity for a ligand and a second effector peptide portion that is
CC chemically or biologically active.They may further comprise a linker
CC peptide between the 2 domains.The oligonucleotides are also designed so
CC that the expressed peptide contains 2 or 4 cysteine residues positioned
CC in, or flanking, the unpredicted or variant residues. These residues
CC confer some degree of conformational rigidity to the peptides. The TSARs
CC or compsns. comprising a TSAR binding domain can be used in vivo to
CC deliver a chemically or biologically active moiety, eg. metal ion,
CC radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC cell. They can also replace the function of macromolecules, eg.
CC monoclonal or polyclonal antibodies and therefore circumvent the need for
CC complex methods of hybridoma formation or in vivo antibody production.
CC The TSARs are easily characterised and have designed activity allowing
CC direct and rapid detection in a screening process.
SQ Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;

Query Match 0.9%; Score 36; DB 12; Length 114;
Best Local Similarity 3.6%; Pred. No. 1.33e-05;
Matches 4; Conservative 34; Mismatches 74; Indels 0; Gaps 0;

```
Db      3 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnnnnnnn 62
          :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :| |  :
Qy     88 CTCCCTGCTGCGCAGCCACTACCGCGAGGTGCTGCCGCTGGCCACGTTCGTGCGGCGCCT 147

       63 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnb 114
          :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :| |  :
Qy    148 GGGGCCCCAGGGCTGGCGGCTGGTGCAGCGCGGGGACCCGGCGGCTTTCCGC 199
```

Search completed: Mon Aug 3 16:47:21 1998
Job time : 1666 secs.

****

MPSRCH (TM)

****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Mon Aug 3 10:56:05 1998; MasPar time 4562.35 Seconds
1232.517 Million cell updates/sec
[redacted]bular output not generated.

Title: >US-08-912-951-1
Description: (1-4015) from US08912951.seq
Perfect Score: 4015
N.A. Sequence: 1 GCAGCGCTGCGTCCTGCTGC.....TTTTTCAGTTTTGAAAAAAA 4015
Comp: CGTCGCGACGCAGGACGACG.....AAAAAGTCAAAACTTTTTTT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 1832099 seqs, 700269816 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: embl-est55
1:em_est10 2:em_est11
Database: genbank-est107
3:gb_est1 4:gb_est10 5:gb_est11 6:gb_est12 7:gb_est13
8:gb_est14 9:gb_est15 10:gb_est16 11:gb_est17
12:gb_est18 13:gb_est19 14:gb_est2 15:gb_est20
16:gb_est21 17:gb_est22 18:gb_est23 19:gb_est3
20:gb_est4 21:gb_est5 22:gb_est6 23:gb_est7 24:gb_est8
25:gb_est9 26:gb_gss

[redacted]istics: Mean 12.222; Variance 2.606; scale 4.689

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	385	9.6	389	8	AA281296	zt08g02.r1 NCI_CGAP_GC	0.00e+00
c	2	315	7.8	340	15	AA811084	oa85c05.s1 NCI_CGAP_GC	0.00e+00
c	3	285	7.1	315	15	AA748707	ny02e05.s1 NCI_CGAP_GC	0.00e+00
	4	248	6.2	409	4	AA311750	EST182469 Jurkat T-cel	0.00e+00
c	5	67	1.7	247	13	AA754458	97SN1784 Rice Immature	2.68e-65
c	6	65	1.6	252	13	AA754459	97SN1787 Rice Immature	5.79e-62
	7	62	1.5	252	13	AA754459	97SN1787 Rice Immature	5.26e-57
	8	51	1.3	247	13	AA754458	97SN1784 Rice Immature	2.34e-39
	9	52	1.3	303	4	AA299878	EST12462 Uterus tumor	6.37e-41
	10	33	0.8	375	25	AA200728	mu13h09.r1 Soares 2NbM	2.86e-13
	11	31	0.8	660	12	AF034177	Homo sapiens ntcon6 co	1.04e-10
	12	33	0.8	2275	12	AF034173	Homo sapiens ntcon2 co	2.86e-13
c	13	33	0.8	2275	12	AF034173	Homo sapiens ntcon2 co	2.86e-13
	14	27	0.7	317	26	AQ008007	CIT-HSP-2288C3.TF CIT-	6.67e-06
	15	27	0.7	560	13	AA775218	ac79e03.s1 Stratagene	6.67e-06
c	16	27	0.7	660	12	AF034177	Homo sapiens ntcon6 co	6.67e-06
	17	28	0.7	1287	13	AF038250	Homo sapiens clone ntc	4.63e-07
c	18	28	0.7	1287	13	AF038250	Homo sapiens clone ntc	4.63e-07
	19	24	0.6	148	10	AA619013	vo66g06.r1 Soares mous	1.25e-02
	20	25	0.6	181	5	AA386387	EST81369 Prostate glan	1.10e-03
	21	25	0.6	187	15	AA855630	vw68h11.r1 Stratagene	1.10e-03
	22	25	0.6	213	14	R72797	yl09c09.r1 Homo sapien	1.10e-03
c	23	24	0.6	275	16	AA899633	UI-R-E0-dg-f-09-0-UI.s	1.25e-02
c	24	25	0.6	278	16	AA882818	TENS0393 T. cruzi epim	1.10e-03
c	25	24	0.6	288	7	C27729	Rice cDNA, partial seq	1.25e-02
c	26	24	0.6	288	24	AA077661	7B35F02 Chromosome 7 F	1.25e-02
	27	25	0.6	301	17	AA848961	EST191723 Normalized r	1.10e-03
	28	25	0.6	318	19	H50134	yo27a07.r1 Homo sapien	1.10e-03
	29	25	0.6	318	17	AA799774	EST189271 Normalized r	1.10e-03
	30	26	0.6	343	23	AA098755	T3985 MVAT4 bloodstrea	8.93e-05
	31	25	0.6	359	4	AA303595	EST16305 Aorta endothe	1.10e-03
	32	25	0.6	382	14	R54656	yj74d04.r1 Homo sapien	1.10e-03
	33	26	0.6	396	20	T26788	T529 Trypanosoma bruce	8.93e-05
	34	26	0.6	400	19	H44707	yp24c08.r1 Homo sapien	8.93e-05
c	35	24	0.6	429	14	R55367	yg88h01.s1 Homo sapien	1.25e-02
c	36	25	0.6	432	5	AA043978	zk58f01.s1 Soares preg	1.10e-03
c	37	25	0.6	438	17	AA924206	UI-R-A1-dx-a-11-0-UI.s	1.10e-03
	38	25	0.6	463	26	AQ000405	CIT-HSP-2288L6.TR CIT-	1.10e-03
c	39	25	0.6	478	17	AA925372	UI-R-A1-ee-d-07-0-UI.s	1.10e-03
	40	25	0.6	487	19	R86860	ym86a03.r1 Homo sapien	1.10e-03
	41	24	0.6	496	10	AA429394	zw32h12.s1 Soares ovar	1.25e-02
c	42	25	0.6	511	18	AA997735	UI-R-C0-hn-b-08-0-UI.s	1.10e-03
c	43	25	0.6	540	16	AA901048	UI-R-A1-dv-e-09-0-UI.s	1.10e-03
	44	25	0.6	590	10	AA622227	no43c05.s1 NCI_CGAP_Pr	1.10e-03
c	45	25	0.6	705	17	AA941159	LD25143.5prime LD Dros	1.10e-03

ALIGNMENTS

RESULT 1
LOCUS AA281296 389 bp mRNA EST 14-AUG-1997
DEFINITION zt08g02.r1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:712562 5', mRNA sequence.
ACCESSION AA281296
NID g1924194
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 389)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 2187 Std Error: 0.00
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 385.
FEATURES Location/Qualifiers
source 1. .389
/organism="Homo sapiens"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTTT-

3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."
/db_xref="taxon:9606"
/clone="IMAGE:712562"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
mRNA <1. .>389
BASE COUNT 87 a 102 c 123 g 77 t
ORIGIN

Query Match 9.6%; Score 385; DB 8; Length 389;
Best Local Similarity 99.5%; Pred. No. 0.00e+00;
Matches 387; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Db      1 GCCAAGTTCCTGCACTGGCTGATGAGTGTGTACGTCGTCGAGCTGCTCAGGTCTTTCTTT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1679 GCCAAGTTCCTGCACTGGCTGATGAGTGTGTACGTCGTCGAGCTGCTCAGGTCTTTCTTT 1738

Db     61 TATGTCACGGAGACCACGTTTCAAAAGAACAGGCTCTTTTTCTACCGGAAGAGTGTCTGG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1739 TATGTCACGGAGACCACGTTTCAAAAGAACAGGCTCTTTTTCTACCGGAAGAGTGTCTGG 1798

Db    121 AGCAAGTTGCAAAGCATTGGAATCAGACAGCACTTGAAGAGGGTGCAGCTGCGGGACGTG 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||  ||
Qy   1799 AGCAAGTTGCAAAGCATTGGAATCAGACAGCACTTGAAGAGGGTGCAGCTGCGGGAGCTG 1858

Db    181 TCGGAAGCAGAGGTCAGGCAGCATCGGGAAGCCAGGCCCGCCCTGCTGACGTCCAGACTC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1859 TCGGAAGCAGAGGTCAGGCAGCATCGGGAAGCCAGGCCCGCCCTGCTGACGTCCAGACTC 1918

Db    241 CGCTTCATCCCCAAGCCTGACGGGCTGCGGCCGATTGTGAACATGGACTACGTCGTGGGA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1919 CGCTTCATCCCCAAGCCTGACGGGCTGCGGCCGATTGTGAACATGGACTACGTCGTGGGA 1978

Db    301 GCCAGAACGTTCCGCAGAGAAAAGAGGGCCGAGCGTCTCACCTCGAGGGTGAAGGCACTG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1979 GCCAGAACGTTCCGCAGAGAAAAGAGGGCCGAGCGTCTCACCTCGAGGGTGAAGGCACTG 2038

Db    361 TTCAGCGTGCTCAACTACGAGCGGGCGCG 389
          |||||||||||||||||||||||||||||
Qy   2039 TTCAGCGTGCTCAACTACGAGCGGGCGCG 2067
```

RESULT 2
LOCUS AA811084 340 bp mRNA EST 19-FEB-1998
DEFINITION oa85c05.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1319048, mRNA sequence.
ACCESSION AA811084
NID g2880695
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 340)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman, Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 2249 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 331.
FEATURES Location/Qualifiers
source 1. .340
/organism="Homo sapiens"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."
/db_xref="taxon:9606"
/clone="IMAGE:1319048"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
BASE COUNT 73 a 86 c 100 g 81 t
ORIGIN

Query Match 7.8%; Score 315; DB 15; Length 340;
Best Local Similarity 99.1%; Pred. No. 0.00e+00;
Matches 328; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

```
Db     12 TTTTTTTCAAAACTGAAAAACTCATATATTCAGTATTTTACTCCCACAGCACCTCCCCCC 71
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   4015 TTTTTTTCAAAACTGAAAAACTCATATATTCAGTATTTTACTCCCACAGCACCTCCCCCC 3956

Db     72 AATTTGACCCACAGG-ACCCCCATCCAGGTGCAGGGTCCTCGCCTGTGTACAGGGCACAC 130
          |||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||
Cp   3955 AATTTGACCCACAGGGACCCCCATCCAGGTGCAGGGTCCTCGCCTGTGTACAGGGCACAC 3896

Db    131 CTTTGGTCACTCCAAATTCCCAGAGCTCCCAGGGTCCTTCTCAGGGTCTCCACCTGGATG 190
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   3895 CTTTGGTCACTCCAAATTCCCAGAGCTCCCAGGGTCCTTCTCAGGGTCTCCACCTGGATG 3836

Db    191 GTTGGGGTGGAAGGCAAAGGAGGGCAGGGCGAGGGGTGAACAATGGCGAATCTGGGGATG 250
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   3835 GTGGGGGTGGAAGGCAAAGGAGGGCAGGGCGAGGGGTGAACAATGGCGAATCTGGGGATG 3776

Db    251 GACTATTCCTATGTGGGGAGTGGAAGCCGGGCTCCTGGTGAGGAAAAGCTGGCCCTGGGG 310
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   3775 GACTATTCCTATGTGGGGAGTGGAAGCCGGGCTCCTGGTGAGGAAAAGCTGGCCCTGGGG 3716

Db    311 TGGAGCCGAGCGCCA-CCTGTGGGGAAGTGA 340
          ||||||||||||||| ||||||||||||||||
Cp   3715 TGGAGCCGAGCGCCAGCCTGTGGGGAAGTGA 3685
```

RESULT 3
LOCUS AA748707 315 bp mRNA EST 18-FEB-1998
DEFINITION ny02e05.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1270592, mRNA sequence.
ACCESSION AA748707
NID g2788665
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 315)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

```
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 1226   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 281.
FEATURES              Location/Qualifiers
     source           1. .315
                      /organism="Homo sapiens"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was prepared from human tonsillar cells enriched for
                      germinal center B cells by flow sorting (CD20+, IgD-),
                      provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                      (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                      primed with a Not I - oligo(dT) primer
                      [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-
                      3']. Double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. Library
                      went through one round of normalization, and was
                      constructed by Bento Soares and M. Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1270592"
                      /clone_lib="NCI_CGAP_GCB1"
                      /tissue_type="germinal center B cell"
                      /lab_host="DH10B"
BASE COUNT       68 a     95 c     87 g     65 t
ORIGIN

  Query Match             7.1%;  Score 285;  DB 15;  Length 315;
  Best Local Similarity  96.2%;  Pred. No. 0.00e+00;
  Matches   302;  Conservative    0;  Mismatches  11;  Indels   1;  Gaps   1;

Db       1 TTTTTTTCAAAACTGAAAAACTCATATATTCAGTATTTTACTCCCACAGCACCTCCCCCC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    4015 TTTTTTTCAAAACTGAAAAACTCATATATTCAGTATTTTACTCCCACAGCACCTCCCCCC 3956

Db      61 AATTTGACCCACAGGGACCCCCATCCAGGTGCAGGGTCCTCGCCTGTGTACAGGGCACAC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    3955 AATTTGACCCACAGGGACCCCCATCCAGGTGCAGGGTCCTCGCCTGTGTACAGGGCACAC 3896

Db     121 CTTTGGTCACTCCAAATTCCCAGAGCTCCCAGGGTCCTTCTCAGGGTCTCCACCTGGATG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    3895 CTTTGGTCACTCCAAATTCCCAGAGCTCCCAGGGTCCTTCTCAGGGTCTCCACCTGGATG 3836

Db     181 GTGGGCCCCGGCCCCCAAACCACGGCACGGCGAGGGGTGAACAATGGCGAATCTGGGGAT 240
          |||||    ||    ||||  | |||| ||||||||||||||||||||||||||||||||
Cp    3835 GTGGGGGT-GGAAGGCAAAGGAGGGCAGGGCGAGGGGTGAACAATGGCGAATCTGGGGAT 3777

Db     241 GGACTATTCCTATGTGGGGAGTGGAAGCCGGGCTCCTGGTGAGGAAAAGCTGGCCCTGGG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    3776 GGACTATTCCTATGTGGGGAGTGGAAGCCGGGCTCCTGGTGAGGAAAAGCTGGCCCTGGG 3717

Db     301 GTGGAGCCGAGCGC 314
          ||||||||||||||
Cp    3716 GTGGAGCCGAGCGC 3703
```

```
RESULT    4
LOCUS       AA311750      409 bp    mRNA            EST       19-APR-1997
DEFINITION  EST182469 Jurkat T-cells VI Homo sapiens cDNA 5' end, mRNA
            sequence.
ACCESSION   AA311750
NID         g1964077
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 409)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
            Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
            Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
            Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES              Location/Qualifiers
     source           1. .409
                      /organism="Homo sapiens"
                      /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                      XhoI"
                      /db_xref="ATCC (inhost):158964"
                      /db_xref="taxon:9606"
                      /clone_lib="Jurkat T-cells VI"
                      /cell_type="T-lymphocyte"
     mRNA             <1. .>409
BASE COUNT       65 a    120 c    133 g     86 t      5 others
ORIGIN

  Query Match             6.2%;  Score 248;  DB 4;  Length 409;
  Best Local Similarity  99.2%;  Pred. No. 0.00e+00;
  Matches   248;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db       1 GTTTGGTGGATGATTTCTTGTTGGTGACACCTCACCTCACCCACGCGAAAACCTTCCTCA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    2649 GTTTGGTGGATGATTTCTTGTTGGTGACACCTCACCTCACCCACGCGAAAACCTTCCTCA 2708

Db      61 GGACCCTGGTCCGAGGTGTCCCTGAGTATGGCTGCGTGGTGAACTTGCGGAAGACAGTGG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    2709 GGACCCTGGTCCGAGGTGTCCCTGAGTATGGCTGCGTGGTGAACTTGCGGAAGACAGTGG 2768

Db     121 TGAACTTCCCTGTAGAAGACGAGGCCCTGGGTGGCACGGCTTTTNTTCAGATGCCGGCCC 180
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
```

```
Qy   2769 TGAACTTCCCTGTAGAAGACGAGGCCCTGGGTGGCACGGCTTTTGTTCAGATGCCGGCCC 2828

Db    181 ACGGCNTATTCCCCTGGTGCGGCCTGCTGCTGGATACCCGGACCCTGGAGGTGCAGAGCG 240
          ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2829 ACGGCCTATTCCCCTGGTGCGGCCTGCTGCTGGATACCCGGACCCTGGAGGTGCAGAGCG 2888

Db    241 ACTACTCCAG 250
          ||||||||||
Qy   2889 ACTACTCCAG 2898


RESULT    5
LOCUS       AA754458      247 bp    mRNA            EST       20-JAN-1998
DEFINITION  97SN1784 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
            cDNA clone 97SN1784, mRNA sequence.
ACCESSION   AA754458
NID         g2801164
KEYWORDS    EST.
SOURCE      rice.
  ORGANISM  Oryza sativa
            Eukaryotae; Viridiplantae; Charophyta/Embryophyta group;
            Embryophyta; Tracheophyta; seed plants; Magnoliophyta; Liliopsida;
            Poales; Poaceae; Oryza.
REFERENCE   1  (bases 1 to 247)
  AUTHORS   Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R, Moon,E.P.,
            Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,
            Lee,M.C. and Eun,M.Y.
  TITLE     Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
  JOURNAL   Unpublished (1998)
COMMENT
            Contact: Eun M.Y.
            Department of Cytogenetics
            National Inst. of Agri. Sci. and Tech, RDA
            Suwon, Kyunggido, Korea
            Tel: 82 331 290 0301
            Fax: 82 331 290 0307
            Email: myeun@sun20.asti.re.kr
            Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
            University, Yongin, Korea. 449-728 bhnahm@bioserver.myongji.ac.kr
            Seq primer: M13 Reverse Primer.
FEATURES              Location/Qualifiers
     source           1. .247
                      /organism="Oryza sativa"
                      /cultivar="Milyang23"
                      /note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
                      XhoI; Directional cDNA library inserted into lambda ZAPII
                      vector  at 5'end with EcoRI and 3' end with Xho I site."
                      /db_xref="taxon:4530"
                      /clone="97SN1784"
                      /clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
                      /tissue_type="Immature Seed"
                      /dev_stage="5 days after pollination"
                      /lab_host="E. coli SOLR"
BASE COUNT        7 a     16 c     21 g     34 t    169 others
ORIGIN

  Query Match             1.7%;  Score 67;  DB 13;  Length 247;
  Best Local Similarity  14.7%;  Pred. No. 2.68e-65;
  Matches    28;  Conservative  101;  Mismatches  58;  Indels   3;  Gaps   3;

Db     55 DCSDNAHCRYTVBWYYARSKYGYGTBYYSWNVDTNTGGTGVGKTTVNVHSGWNNRCSNSV 114
          :| : |:|:  :::: :::: :||::::: ::  | | |:|   : :: |:   |  ::
Cp   1406 GCTGCACCAGGCGACGGGGGTCTGTGTCCTCCTCCTCGGGGGCCGCCACAGAGCCCTGGG 1347

Db    115 VYVWBTAYCDYBHYBDRANHV-DD-TRCTNDRGYCNYTASDNGTSATKRVTG-YDKTDSD 171
          :: ::| :|: :: :::| :: :: |:|| ::|:  :  ::  |:   :: | :::|:::
Cp   1346 GCTTCTCCCGGGCACAGACACCGGCTGCTGGGGTGACCGCAGCTCGCAGCGGGCAGTGCG 1287

Db    172 CGGGCWRKVTYGSSBYBRCGVNVMVRTTSMWTDKSTKMBSMDMSRRSRVHYGRWMBNKKR 231
            :::: : :::::  |: ::: | :   :::|:::::: :::: :: |:::: ::::
Cp   1286 TCTTGAGGAGCACCCCGTAGGGGCACTGCGCGTGGTTCCCAAGCAGCTCCAGAAACAGGG 1227

Db    232 GMSRNWTDTK 241
```

```
          |::: :|:|:
Cp   1226 GCCGCATTTG 1217


RESULT    6
LOCUS       AA754459      252 bp    mRNA            EST       20-JAN-1998
DEFINITION  97SN1787 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
            cDNA clone 97SN1787, mRNA sequence.
ACCESSION   AA754459
NID         g2801165
KEYWORDS    EST.
SOURCE      rice.
  ORGANISM  Oryza sativa
            Eukaryotae; Viridiplantae; Charophyta/Embryophyta group;
            Embryophyta; Tracheophyta; seed plants; Magnoliophyta; Liliopsida;
            Poales; Poaceae; Oryza.
REFERENCE   1  (bases 1 to 252)
  AUTHORS   Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R, Moon,E.P.,
            Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,
            Lee,M.C. and Eun,M.Y.
  TITLE     Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
  JOURNAL   Unpublished (1998)
COMMENT
            Contact: Eun M.Y.
            Department of Cytogenetics
            National Inst. of Agri. Sci. and Tech, RDA
            Suwon, Kyunggido, Korea
            Tel: 82 331 290 0301
            Fax: 82 331 290 0307
            Email: myeun@sun20.asti.re.kr
            Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
            University, Yongin, Korea. 449-728 bhnahm@bioserver.myongji.ac.kr
            Seq primer: M13 Reverse Primer.
FEATURES              Location/Qualifiers
     source           1. .252
                      /organism="Oryza sativa"
                      /cultivar="Milyang23"
                      /note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
                      XhoI; Directional cDNA library inserted into lambda ZAPII
                      vector  at 5'end with EcoRI and 3' end with Xho I site."
                      /db_xref="taxon:4530"
                      /clone="97SN1787"
                      /clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
                      /tissue_type="Immature Seed"
                      /dev_stage="5 days after pollination"
                      /lab_host="E. coli SOLR"
BASE COUNT        5 a     21 c     12 g     35 t    179 others
ORIGIN

  Query Match             1.6%;  Score 65;  DB 13;  Length 252;
  Best Local Similarity  12.4%;  Pred. No. 5.79e-62;
  Matches    27;  Conservative  114;  Mismatches  74;  Indels   3;  Gaps   2;

Db     21 TSYBCHGNBVWVCVASHGNYMSVHNCTBRGTHCDCKNVNWSTMTW-GTVNWBNVSGDWHY 79
          |::: :  :  :|: ::  ::::: |   | :|:|: : :: :   ||: :: ::|: :
Cp    886 TCTGGCAGGTGACACCACACAGAAACCACGGTCACTCGGTCCACGCGTCCTGCCCGGGTG 827

Db     80 WBVBNTKVDVGNHTRCSRWRBVTRMAHYHDYTNCBBYNNNDYHMWHBBMYBBTGCMTCTM 139
           :::  ::::  :|:|: :::: :: :::::| |:::   :    ::: ::  ||:|
Cp    826 GGCCCAGGACCCCTGCCCAACGGGCGTCCGCTCCGGCTCAGGGGCAGCGCCACGCCTGGG 767

Db    140 WCWBHYNTKCTASGWHTSTNYDVKSSTNTWGVTBSYDKSMHGYWCSBBVKYHTKVSTTRA 199
           |:::  :| |:| : :  :::::::|    :|:::  : :|:  :::::::  ::  :
Cp    766 CCTCTTGGGCAACGGCAGACTTCGGCTGGCACTGCCCCCGCGCCTCCTCGCACCCGGGGC 707

Db    200 TRSYTCVRKYCVMWMTKKVVKKYHVVBBGCHBTDSKCK 237
          |:::  ::::|::   ::::  : :::: |:: :::|:
Cp    706 TGGC--AGGCCCAGGGGGACCCCGGCCTCCCTGACGCT 671


RESULT    7
LOCUS       AA754459      252 bp    mRNA            EST       20-JAN-1998
DEFINITION  97SN1787 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
```

```
            cDNA clone 97SN1787, mRNA sequence.
ACCESSION   AA754459
NID         g2801165
KEYWORDS    EST.
SOURCE      rice.
  ORGANISM  Oryza sativa
            Eukaryotae; Viridiplantae; Charophyta/Embryophyta group;
            Embryophyta; Tracheophyta; seed plants; Magnoliophyta; Liliopsida;
            Poales; Poaceae; Oryza.
REFERENCE   1  (bases 1 to 252)
  AUTHORS   Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R, Moon,E.P.,
            Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,
            Lee,M.C. and Eun,M.Y.
  TITLE     Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
  JOURNAL   Unpublished (1998)
COMMENT
            Contact: Eun M.Y.
            Department of Cytogenetics
            National Inst. of Agri. Sci. and Tech, RDA
            Suwon, Kyunggido, Korea
            Tel: 82 331 290 0301
            Fax: 82 331 290 0307
            Email: myeun@sun20.asti.re.kr
            Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
            University, Yongin, Korea. 449-728 bhnahm@bioserver.myongji.ac.kr
            Seq primer: M13 Reverse Primer.
FEATURES              Location/Qualifiers
     source           1. .252
                      /organism="Oryza sativa"
                      /cultivar="Milyang23"
                      /note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
                      XhoI; Directional cDNA library inserted into lambda ZAPII
                      vector  at 5'end with EcoRI and 3' end with Xho I site."
                      /db_xref="taxon:4530"
                      /clone="97SN1787"
                      /clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
                      /tissue_type="Immature Seed"
                      /dev_stage="5 days after pollination"
                      /lab_host="E. coli SOLR"
BASE COUNT        5 a     21 c     12 g     35 t    179 others
ORIGIN

  Query Match             1.5%;  Score 62;  DB 13;  Length 252;
  Best Local Similarity  11.7%;  Pred. No. 5.26e-57;
  Matches    27;  Conservative  118;  Mismatches  82;  Indels   3;  Gaps   3;

Db      23 YBCHGNBVWVCVASHGNYMSVHNCTBRGTHCDCKNVNWSTMTWGTVNWBNVSGDWHYWBV 82
           ::|:| ::::|: ::  ::::: |: | :|    :    : :|    : : | : ::::
Qy    3021 TGCAGGTGAACAGCCTCCAGACGGTGTGCACCAACATCTACAAGATCCTCCTGCTGCAGG 3080

Db      83 BNTKVDVGNHT-RCS-RWRBVTRMAHYHDYTNCBBYNNNDYHMWHBBMYBBTGCMTCTMW 140
           : | :::| :| :|: :::::|::  ::::| |: :   :::: :  : ::|| :   :
Qy    3081 CGTACAGGTTTCACGCATGTGTGCTGCAGCTCCCATTTCATCAGCA-AGTTTGGAAGAAC 3139

Db     141 CWBHYNTKCTASGWHTSTNYDVKSSTNTWGVTBSYDKSMHGYWCSBBVKYHTKVSTTRAT 200
           | ::: |: | : :  :  : :::    : : ::: :::: :: :: :::: :::| :|
Qy    3140 CCCACATTTTTCCTGCGCGTCATCTCTGACACGGCCTCCCTCTGCTACTCCATCCTGAAA 3199

Db     201 RSYTCVRKYCVMWMTKKVVKKYHVVBBGCHBTDSKCKTMWMTNKHVMTST 250
           :::  :: : ::    ::  ::::::::||:  :::|: : :  ::::|:|
Qy    3200 GCCAAGAACGCAGGGATGTCGCTGGGGGCCAAGGGCGCCGCCGGCCCTCT 3249


RESULT    8
LOCUS       AA754458      247 bp    mRNA            EST       20-JAN-1998
DEFINITION  97SN1784 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
            cDNA clone 97SN1784, mRNA sequence.
ACCESSION   AA754458
NID         g2801164
KEYWORDS    EST.
SOURCE      rice.
  ORGANISM  Oryza sativa
            Eukaryotae; Viridiplantae; Charophyta/Embryophyta group;
            Embryophyta; Tracheophyta; seed plants; Magnoliophyta; Liliopsida;
            Poales; Poaceae; Oryza.
REFERENCE   1  (bases 1 to 247)
  AUTHORS   Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R, Moon,E.P.,
            Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,
            Lee,M.C. and Eun,M.Y.
  TITLE     Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
  JOURNAL   Unpublished (1998)
COMMENT
            Contact: Eun M.Y.
            Department of Cytogenetics
            National Inst. of Agri. Sci. and Tech, RDA
            Suwon, Kyunggido, Korea
            Tel: 82 331 290 0301
            Fax: 82 331 290 0307
            Email: myeun@sun20.asti.re.kr
            Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
            University, Yongin, Korea. 449-728 bhnahm@bioserver.myongji.ac.kr
            Seq primer: M13 Reverse Primer.
FEATURES              Location/Qualifiers
     source           1. .247
                      /organism="Oryza sativa"
                      /cultivar="Milyang23"
                      /note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
                      XhoI; Directional cDNA library inserted into lambda ZAPII
                      vector  at 5'end with EcoRI and 3' end with Xho I site."
                      /db_xref="taxon:4530"
                      /clone="97SN1784"
                      /clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
                      /tissue_type="Immature Seed"
                      /dev_stage="5 days after pollination"
                      /lab_host="E. coli SOLR"
BASE COUNT        7 a     16 c     21 g     34 t    169 others
ORIGIN

  Query Match             1.3%;  Score 51;  DB 13;  Length 247;
  Best Local Similarity  10.5%;  Pred. No. 2.34e-39;
  Matches    23;  Conservative  108;  Mismatches  86;  Indels   2;  Gaps   2;

Db      17 BAWMNKHTHMMTBBWCCVRRVGTTTNNGKHNGRTTTWNDCSDNAHCRYTVBWYYARSKYG 76
           :|:: :  ::: :::|| : : |    | :  :| |  :|:: :|:  :: :: ::::
Qy    1672 GATCCTGGCCAAGTTCCTGCACTGGCTGATGAGTGTGTACGTCGTCGAGCTGCTCAGGTC 1731

Db      77 YGTBYYSWNVDTNTGGTGVGKTTVNVHSGWNNRCSNSVVYVWBTAYCDYBHYBDRANHVD 136
           : |::: :  :|   | :    :  :  :  :   ::: :::| : :::::::::|  ::
Qy    1732 TTTCTTTTATGTCACGGAGACCACGTTTCAAAAGAACAGGCTCTTTTTCTACCGGAAGAG 1791

Db     137 DTRCTNDRGYCN-YTASDNGTSATKRVTGYDKTDSDCGGGCWRKVTYGSSBYBRCGVNVM 195
           :  || ::|:   :| ::   :||::: :   ::: |  |: ::   ::::: |: ::
Qy    1792 TGTCTGGAGCAAGTTGCAAAGCATTGGAATCAGACA-GCACTTGAAGAGGGTGCAGCTGC 1850

Db     196 VRTTSMWTDKSTKMBSMDMSRRSRVHYGRWMBNKKRGMS 234
           ::  ::: : :  :::::::: ::: : : :: ::: ::
Qy    1851 GGGAGCTGTCGGAAGCAGAGGTCAGGCAGCATCGGGAAG 1889


RESULT    9
LOCUS       AA299878      303 bp    mRNA            EST       18-APR-1997
DEFINITION  EST12462 Uterus tumor I Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION   AA299878
NID         g1952209
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 303)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
```

```
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
            Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
            Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
            Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES              Location/Qualifiers
     source           1. .303
                      /organism="Homo sapiens"
                      /note="Organ: uterus; Vector: pBluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI"
                      /db_xref="ATCC (inhost):193066"
                      /db_xref="taxon:9606"
                      /clone_lib="Uterus tumor I"
                      /dev_stage="adult"
     mRNA             <1. .>303
BASE COUNT       54 a     85 c     93 g     69 t      2 others
ORIGIN

  Query Match             1.3%;  Score 52;  DB 4;  Length 303;
  Best Local Similarity  98.1%;  Pred. No. 6.37e-41;
  Matches    53;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db       1 CTGACACGGCCTCCCTCTGCTACTCCATCCTGAAAGCCAAGAACGCAGGTATGT 54
           ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Qy    3165 CTGACACGGCCTCCCTCTGCTACTCCATCCTGAAAGCCAAGAACGCAGGGATGT 3218


RESULT   10
LOCUS       AA200728      375 bp    mRNA            EST       19-FEB-1997
DEFINITION  mu13h09.r1 Soares 2NbMT Mus musculus cDNA clone 639329 5', mRNA
            sequence.
ACCESSION   AA200728
NID         g1795738
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 375)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:391321
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 361.
FEATURES              Location/Qualifiers
     source           1. .375
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCGTTTTTTTTTTTTTTTTTTTTTTTT
                      3']; double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. RNA
                      provided by Dr. Bertrand Jordan. Library went through two
                      rounds of normalization, and was constructed by Bento
                      Soares and M.Fatima Bonaldo."
                      /db_xref="taxon:10090"
                      /clone="639329"
                      /clone_lib="Soares 2NbMT"
                      /sex="male"
                      /tissue_type="Thymus"
                      /dev_stage="4 weeks"
                      /lab_host="DH10B"
     mRNA             <1. .>375
BASE COUNT       80 a     95 c    105 g     95 t
ORIGIN

  Query Match             0.8%;  Score 33;  DB 25;  Length 375;
  Best Local Similarity  75.4%;  Pred. No. 2.86e-13;
  Matches    49;  Conservative    0;  Mismatches  16;  Indels   0;  Gaps   0;

Db     144 AGAGCATCTCTATGAATGAGAGCAGCAGCAGCCTGTTTGACTTCTTCCTGCACTTCCTGC 203
           |||||  |||  ||||||||  ||||||  |||| || ||| |||||||  | |||| ||
Qy    2436 AGAGCTCCTCCCTGAATGAGGCCAGCAGTGGCCTCTTCGACGTCTTCCTACGCTTCATGT 2495

Db     204 GTCAC 208
           | |||
Qy    2496 GCCAC 2500


RESULT   11
LOCUS       AF034177      660 bp    mRNA            EST       22-DEC-1997
DEFINITION  Homo sapiens ntcon6 contig mRNA, partial sequence, mRNA sequence.
ACCESSION   AF034177
NID         g2707739
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 660)
  AUTHORS   Tripodis,N. and Ragoussis,J.
  TITLE     Generation of a transcription map in the region immediately
            centromeric to human MHC across the 6p21.2-6p21.3 chromosomal
            boundary
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 660)
  AUTHORS   Tripodis,N. and Ragoussis,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-1997) Division of Medical and Molecular Genetics,
            Guy's Hospital, 7th floor, Guy's Tower, London SE1 9RT, UK
FEATURES              Location/Qualifiers
     source           1. .660
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
```

```
                     /chromosome="6"
                     /map="6p21.3"
                     /clone="ntcon6 contig"
                     /tissue_type="fetal brain; fetal liver; adult muscle"
                     /note="similar to CutA"
BASE COUNT       162 a    161 c    173 g    109 t     55 others
ORIGIN

  Query Match             0.8%;  Score 31;  DB 12;  Length 660;
  Best Local Similarity  15.6%;  Pred. No. 1.04e-10;
  Matches     7;  Conservative   31;  Mismatches   7;  Indels   0;  Gaps   0;

Db     384 WSSRCCYRSYGMYHYYYCYYKKCDMSCTTSKSRWYKSKRSWCDCD 428
           :: :| :::::| ::::::  :::|:::|| :::::::::::: |:|:
Qy     336 AGAACGTGCTGGCCTTCGGCTTCGCGCTGCTGGACGGGGCCCGCG 380


RESULT   12
LOCUS       AF034173     2275 bp    mRNA            EST       22-DEC-1997
DEFINITION  Homo sapiens ntcon2 contig mRNA, partial sequence, mRNA sequence.
ACCESSION   AF034173
NID         g2707735
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2275)
  AUTHORS   Tripodis,N. and Ragoussis,J.
  TITLE     Generation of a transcription map in the region immediately
            centromeric to human MHC across the 6p21.2-6p21.3 chromosomal
            boundary
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2275)
  AUTHORS   Tripodis,N. and Ragoussis,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-1997) Division of Medical and Molecular Genetics,
            Guy's Hospital, 7th floor, Guy's Tower, London SE1 9RT, UK
FEATURES             Location/Qualifiers
     source          1. .2275
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6p21.3"
                     /clone="ntcon2 contig"
                     /tissue_type="liver; brain"
                     /dev_stage="fetus"
                     /note="similar to Br140"
BASE COUNT       438 a    619 c    470 g    599 t    149 others
ORIGIN

  Query Match             0.8%;  Score 33;  DB 12;  Length 2275;
  Best Local Similarity  13.3%;  Pred. No. 2.86e-13;
  Matches    11;  Conservative   47;  Mismatches  25;  Indels   0;  Gaps   0;

Db    1523 RGMKKCWKYRKYKKYTSTYYKSWSRWYWYTTYTYWYCWCCTSMKSASCAMMRWMGYMGSR 1582
           :|:::|: : ::::::|: : : :::::::|| |:     | ::::|:| ::  : :  ::
Qy    2637 GGCTGCTCCTGCGTTTGGTGGATGATTTCTTGTTGGTGACACCTCACCTCACCCACGCGA 2696

Db    1583 SSRSYWGYWGSMSGCYGMTKRYY 1605
             :::: :  :::| : :|::::
Qy    2697 AAACCTTCCTCAGGACCCTGGTC 2719


RESULT   13
LOCUS       AF034173     2275 bp    mRNA            EST       22-DEC-1997
DEFINITION  Homo sapiens ntcon2 contig mRNA, partial sequence, mRNA sequence.
ACCESSION   AF034173
NID         g2707735
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2275)
  AUTHORS   Tripodis,N. and Ragoussis,J.
  TITLE     Generation of a transcription map in the region immediately
            centromeric to human MHC across the 6p21.2-6p21.3 chromosomal
            boundary
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2275)
  AUTHORS   Tripodis,N. and Ragoussis,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-1997) Division of Medical and Molecular Genetics,
            Guy's Hospital, 7th floor, Guy's Tower, London SE1 9RT, UK
FEATURES             Location/Qualifiers
     source          1. .2275
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6p21.3"
                     /clone="ntcon2 contig"
                     /tissue_type="liver; brain"
                     /dev_stage="fetus"
                     /note="similar to Br140"
BASE COUNT       438 a    619 c    470 g    599 t    149 others
ORIGIN

  Query Match             0.8%;  Score 33;  DB 12;  Length 2275;
  Best Local Similarity  13.0%;  Pred. No. 2.86e-13;
  Matches    15;  Conservative   64;  Mismatches  34;  Indels   2;  Gaps   2;

Db    1483 RWKRRKGRRKRMTGMYKRMYRAMMMAMCAMMACWWYYWKMRGMKKCWKYRKYKKYTS-TY 1541
           ::::::|:::   | :: :   :::|: |:: |::::: :: : :|:: ::: :: : |:
Cp    2692 GTGGGTGAGGTGAGGTGTCACCAACAAGAAATCATCCACCAAACGCAGGAGCAGCCCGTC 2633

Db    1542 YKSWSRWYWYTTYTYWY-CWCCTSMKSASCAMMRWMGYMGSRSSRSYWGYWGSMS 1595
           : : ::: ::  : ::  |: ||  :   | : :::|::| ::::: |:::|:::
Cp    2632 CCGCCGAATCCCCGCAAACAGCTTGTTCTCCATGTCGCCGTAGCACAGGCTGCAG 2578


RESULT   14
LOCUS       AQ008007     317 bp    DNA             GSS       24-APR-1998
DEFINITION  CIT-HSP-2288C3.TF CIT-HSP Homo sapiens genomic clone 2288C3,
            genomic survey sequence.
ACCESSION   AQ008007
NID         g3085233
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 317)
  AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
            Simon,M. and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13-21;
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .317
                     /organism="Homo sapiens"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
```

```
                     /db_xref="taxon:9606"
                     /clone="2288C3"
                     /clone_lib="CIT-HSP"
                     /sex="Male"
                     /cell_type="Sperm"
BASE COUNT       59 a     92 c     91 g     75 t
ORIGIN

  Query Match            0.7%;  Score 27;  DB 26;  Length 317;
  Best Local Similarity  86.5%;  Pred. No. 6.67e-06;
  Matches    32;  Conservative    0;  Mismatches   5;  Indels   0;  Gaps   0;

Db    231 CCTTTCCCCAGGTGTCCTGTCTGAAGGAGATGGGGGC 267
          |||| | |||||||||||| ||||||||| ||| |||
Qy    264 CCTTCCGCCAGGTGTCCTGCCTGAAGGAGCTGGTGGC 300


RESULT   15
LOCUS       AA775218      560 bp    mRNA            EST       05-FEB-1998
DEFINITION  ac79e03.s1 Stratagene lung (#937210) Homo sapiens cDNA clone 868828
            3' similar to gb:J00194 HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, DR
            ALPHA CHAIN (HUMAN);, mRNA sequence.
ACCESSION   AA775218
NID         g2834552
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 560)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 344.
FEATURES             Location/Qualifiers
     source          1. .560
                     /organism="Homo sapiens"
                     /note="Organ: lung; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. normal lung.  Average insert size: 1.0 kb;
                     Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCACGAG
                     3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
                     /db_xref="taxon:9606"
                     /clone="868828"
                     /clone_lib="Stratagene lung (#937210)"
                     /sex="male"
                     /dev_stage="72 years"
                     /lab_host="SOLR cells (kanamycin resistant)"
BASE COUNT      173 a    118 c    134 g    135 t
ORIGIN

  Query Match            0.7%;  Score 27;  DB 13;  Length 560;
  Best Local Similarity  82.9%;  Pred. No. 6.67e-06;
  Matches    34;  Conservative    0;  Mismatches   7;  Indels   0;  Gaps   0;

Db    466 TGATGCCCACCAGACCACAGTCAGGCCCAGGGCACACACCA 506
          ||||| ||||| | |||||| |||||| || ||| ||||||
Qy   3452 TGATGGCCACCCGCCCACAGCCAGGCCGAGAGCAGACACCA 3492


Search completed: Mon Aug  3 13:21:09 1998
Job time : 8704 secs.
```